SEARCH REQUEST FORM

Scientific and Technical Information Center 3817

Requester's Full Name _____ Examiner # :_____ Date: _____
Art Unit: _____ Phone Number 30_____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

****

STAFF USE ONLY

Searcher Shelly Gossard
Searcher Phone # 308-4494
Searcher Location _____
Date Searcher Picked Up _____
Date Completed 6/5/00
Searcher Prep & Review Time _____
Clerical Prep Time _____
Online Time _____

PTO-1590 (2000)

Type of Search

NA Sequence (#)_____
AA Sequence (#)_____
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis _____
Sequence Systems _____
WWW/Internet _____
Other (specify) _____

```
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1069 AA;  116314 MW;  0F3F60C6 CRC32;

  Query Match           10.8%;  Score 96;  DB 11;  Length 1069;
  Best Local Similarity  50.0%;  Pred. No. 2.17e-02;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    887 GIMTVILIILIVMMARYCRSKNKNGYEA 914
          ||:|||| :|:::  :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT   14
ID   O04557      PRELIMINARY;      PRT;   614 AA.
AC   O04557;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   T7N9.10.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BUEHLER E., DEWAR K., FENG J., KIM C., LI Y., SHINN P., SUN H.,
RA   CONWAY A., CONWAY A., KURTZ D., OJI O., OSBORNE B., SHEN Y.K.,
RA   TORIUMI M., VYSOTSKAIA V., YU G., DAVIS R.W., FEDERSPIEL N.A.,
RA   THEOLOGIS A., ECKER J.R.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC000348; AAB61486.1; -.
DR   MENDEL; 16818; Arath;2604;16818.
DR   PFAM; PF00854; PTR2; 1.
SQ   SEQUENCE   614 AA;  69463 MW;  7D875C1F CRC32;

  Query Match           10.7%;  Score 95;  DB 10;  Length 614;
  Best Local Similarity  45.8%;  Pred. No. 3.19e-02;
  Matches    11;  Conservative    5;  Mismatches   8;  Indels   0;  Gaps   0;

Db    564 LIAVLGVVNLVYFWYCAHRYQYKA 587
          |  :|||: |:  ||| :|  |:|
Qy     33 LTVILGVLLLIGCWYCRRRNGYRA 56


RESULT   15
ID   O61802      PRELIMINARY;      PRT;  1510 AA.
AC   O61802;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   H11E01.3 PROTEIN.
GN   H11E01.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   GEISEL C., WAMSLEY P.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF067222; AAC17017.1; -.
SQ   SEQUENCE   1510 AA;  166889 MW;  C26457A9 CRC32;

  Query Match           10.6%;  Score 94;  DB 5;  Length 1510;
  Best Local Similarity  33.3%;  Pred. No. 4.67e-02;
  Matches    13;  Conservative   13;  Mismatches  12;  Indels   1;  Gaps   1;

Db    132 RRYS-NSFEDRDFVQRIEDRNTDPVEQNTKDPYKRLSAN 169
          || : ::|: ||    ::::| :||  |:  :| :|||:
Qy     72 RRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAE 110


Search completed: Fri May  5 21:51:04 2000
Job time : 184 secs.
```

```
Db     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT    2
ID   R84212 standard; Protein; 118 AA.
AC   R84212;
DT   20-APR-1996  (first entry)
DE   MART-1 melanoma antigen.
KW   MART-1; melanoma antigen recognised by T-cell; melanoma;
KW   metastatic melanoma; tumour-associated antigen; immunogen;
KW   diagnosis; prognosis; prophylaxis; therapy; vaccine.
OS   Mammalian.
FH   Key             Location/Qualifiers
FT   region          27. .47
FT                   /note= "hydrophobic region"
PN   WO9529193-A2.
PD   02-NOV-1995.
PF   21-APR-1995; U05063.
PR   22-APR-1994; US-231565.
PR   05-APR-1995; US-417174.
PA   (USSH ) US SEC DEPT HEALTH.
PI   Kawakami Y,  Rosenberg SA;
DR   WPI; 95-382963/49.
DR   N-PSDB; T02714.
PT   DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT   vectors, host cells and antibodies, used to detect, treat and
PT   immunise animal against melanoma.
PS   Claim 11; Page 117; 184pp; English.
CC   The melanoma antigen (MART-1) is produced by recombinant DNA
CC   methods, i.e. preferably using a baculovirus vector for expression
CC   in insect cell cultures.  MART-1 protein is a source of immunogenic
CC   peptides (see R84196 for peptide M9-2) which are optionally modified
CC   (see R84783-R84800) and used in medicaments for the treatment or
CC   prevention (by immunization) of melanoma.  Antibodies against MART-1
CC   and its immunogenic peptides may be used in the detection and
CC   isolation of MART-1 from a sample, the detection of which is
CC   indicative of a disease state (melanoma or metastatic melanoma).
SQ   Sequence   118 AA;

  Query Match          100.0%;  Score 889;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 1.61e-82;
  Matches   118;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Db     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT    3
ID   R63158 standard; Protein; 118 AA.
AC   R63158;
DT   26-MAY-1995  (first entry)
DE   Tumour rejection antigen precursor.
KW   Tumour rejection antigen; precursor; HLA-A2 molecule; tyrosinase;
KW   isolation; melanoma; cell line; LB-39-MEL; diagnosis; vaccine;
KW   therapy.
OS   Homo sapiens.
PN   WO9421126-A.
PD   29-SEP-1994.
PF   09-MAR-1994; U02487.
PR   18-MAR-1993; US-032978.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon-Falleur T,  Brichard V,  De Plaen E,  Traversari C;
PI   Van Pel A, Wolfel T;
DR   WPI; 94-316544/39.
DR   N-PSDB; Q76370.
PT   Nucleic acid coding for a tumour rejection antigen precursor - is
PT   used for developing prods. for diagnosis or treatment of expression
PT   related disorders, partic. melanoma
PS   Claim 5; Page 14; 26pp; English.
CC   This sequence represents the tumour rejection antigen precursor which is
CC   processed to a tumour rejection antigen presented by HLA-A2 molecules.
CC   The tumour rejection antigen is not related to tyrosinase.  The cDNA
CC   encoding this sequence was isolated from the melanoma cell line,
CC   LB-39-MEL.  The tumour rejection antigen may be used for diagnosis or
CC   in vaccines or for therapy of disorders characterised by the expression
CC   of the tumour rejection antigen precursor, particularly melanoma.
SQ   Sequence   118 AA;

  Query Match          100.0%;  Score 889;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 1.61e-82;
  Matches   118;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Db     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT    4
ID   W00903 standard; Peptide; 21 AA.
AC   W00903;
DT   23-MAY-1997  (first entry)
DE   Human melanoma MART-1/Aa tumour associated antigen p27-47.
KW   Adeno-associated virus; vector; liposome; transfection;
KW   dendritic cell; melanoma; MART-1/Aa; adoptive immunotherapy;
KW   tumour associated antigen.
OS   Homo sapiens.
PN   WO9703703-A1.
PD   06-FEB-1997.
PF   19-JUL-1996; U12012.
PR   21-JUL-1995; US-001312.
PR   01-NOV-1995; US-007184.
PR   01-DEC-1995; US-566286.
PA   (RHON ) RHONE POULENC RORER PHARM INC.
PI   Lebkowski JS,  Philip R;
DR   WPI; 97-145208/13.
PT   Adeno-associated virus:liposome complexes for transfecting dendritic
PT   cells - for inducing immune response, useful for treating e.g.
PT   neoplasia or infections
PS   Example 5; Page 58; 134pp; English.
CC   Tumour associated antigens (W13660-61, W00878-903) can be loaded
CC   into dendritic cells and used to induce antitumour immunity.
CC   Alternatively, the dendritic cells are transfected with adeno
CC   associated virus plasmid DNA (which includes DNA encoding the
CC   tumour associated antigen) complexed with cationic liposomes.  The
CC   amtigen loaded or transfected dendritic cells can be used to
CC   generate tumour antigen-specific cytotoxic T lymphocytes for use in
CC   adoptive immunotherapy in a patient having the corresponding
CC   tumour.  A suitable antigen comprises amino acids 27-47 (W00903)
CC   of human melanoma MART-1/Aa.
SQ   Sequence   21 AA;

  Query Match          18.7%;  Score 166;  DB 1;  Length 21;
  Best Local Similarity 100.0%;  Pred. No. 1.81e-06;
  Matches    21;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTVILGVLLLIGCWY 21
          |||||||||||||||||||||
Qy     27 AAGIGILTVILGVLLLIGCWY 47


RESULT    5
ID   W63682 standard; Protein; 291 AA.
AC   W63682;
DT   24-SEP-1998  (first entry)
DE   Human secreted protein 2.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 21:43:29 2000; MasPar time 5.72 Seconds
488.454 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-2
Description: (1-118) from US09267439.pep
Perfect Score: 889
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 29.849; Variance 115.902; scale 0.258

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	889	100.0	118	1	W83134	Human tumour rejection	1.61e-82
2	889	100.0	118	1	R84212	MART-1 melanoma antige	1.61e-82
3	889	100.0	118	1	R63158	Tumour rejection antig	1.61e-82
4	166	18.7	21	1	W00903	Human melanoma MART-1/	1.81e-06
5	94	10.6	291	1	W63682	Human secreted protein	5.23e+00
6	90	10.1	380	1	R05433	cPA-P2 Hybrid plasmino	1.11e+01
7	89	10.0	226	1	W80407	A secreted protein enc	1.34e+01
8	88	9.9	141	1	W20924	H. pylori cell envelop	1.61e+01
9	88	9.9	169	1	Y11007	H. pylori ORF hp6p1050	1.61e+01
10	88	9.9	215	1	Y11016	H. pylori ORF hp6p1050	1.61e+01
11	87	9.8	519	1	W30826	The novel tyrosinase-r	1.94e+01
12	84	9.4	296	1	W20802	H. pylori inner membra	3.37e+01
13	83	9.3	200	1	R82900	Mouse B7-1 (IgV-like d	4.04e+01
14	83	9.3	212	1	R82902	Mouse B7-1 IgV-like is	4.04e+01
15	83	9.3	306	1	R67990	Murine B lymphocyte an	4.04e+01
16	83	9.3	306	1	R82893	Mouse B7-1 alternative	4.04e+01
17	83	9.3	306	1	W73641	Mouse B7-2 antigen.	4.04e+01
18	83	9.3	306	1	W67805	Mouse B7 protein seque	4.04e+01
19	82	9.2	433	1	W76411	Human betac cytoplasmi	4.85e+01
20	82	9.2	641	1	W35298	Macaque islet cell ant	4.85e+01
21	82	9.2	818	1	W35297	Human islet cell antig	4.85e+01
22	82	9.2	876	1	W18091	Type I diabetes-associ	4.85e+01
23	82	9.2	897	1	R20982	Sequence of beta-chain	4.85e+01
24	82	9.2	969	1	W25170	Human insulinoma-assoc	4.85e+01
25	82	9.2	986	1	W25171	Human insulinoma-assoc	4.85e+01
26	82	9.2	1012	1	W35296	Macaque islet cell ant	4.85e+01
27	82	9.2	1015	1	W35345	Human protein tyrosine	4.85e+01
28	82	9.2	1015	1	W18092	Type I diabetes-associ	4.85e+01
29	81	9.1	455	1	R28757	Hepatocyte nuclear fac	5.80e+01
30	81	9.1	852	1	W79065	FIV-NCSU1 clone JSY3 e	5.80e+01
31	81	9.1	855	1	W23711	Feline immunodeficienc	5.80e+01
32	81	9.1	855	1	R51252	FIV UK2 envelope prote	5.80e+01
33	81	9.1	856	1	R51248	FIV petaluma envelope	5.80e+01
34	81	9.1	856	1	W08436	Feline immunodeficienc	5.80e+01
35	81	9.1	856	1	W01823	FIV envelope polypepti	5.80e+01
36	81	9.1	856	1	R51247	FIV envelope protein c	5.80e+01
37	81	9.1	856	1	W07846	FIV envelope protein.	5.80e+01
38	81	9.1	856	1	R51249	FIV PET-F14 envelope p	5.80e+01
39	81	9.1	856	1	W53671	FIV PPR clone 34 ORF5	5.80e+01
40	81	9.1	1978	1	Y07032	Breast cancer associat	5.80e+01
41	80	9.0	522	1	W73047	Mokola virus G glycopr	6.94e+01
42	79	8.9	163	1	Y12385	Human 5' EST secreted	8.30e+01
43	79	8.9	234	1	R13844	Fusaric acid resistanc	8.30e+01
44	79	8.9	279	1	W74908	Human secreted protein	8.30e+01
45	79	8.9	854	1	R51250	FIV PPR envelope prote	8.30e+01

ALIGNMENTS

RESULT 1
ID W83134 standard; Protein; 118 AA.
AC W83134;
DT 04-FEB-1999 (first entry)
DE Human tumour rejection antigen precursor.
KW Human; tumour rejection antigen precursor; human leukocyte antigen;
KW TRAP; HLA; cancer; melanoma.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 2
FT /note= "encoded by CGA"
FT Misc_difference 17
FT /note= "encoded by GAC"
PN US5837476-A.
PD 17-NOV-1998.
PF 16-JAN-1998; 007966.
PR 03-MAR-1995; US-398409.
PR 16-JAN-1998; US-007966.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, Brichard V, De Plaen E, Traversari C,
PI Van Pel A, Woelfelt;
DR WPI; 99-043967/04.
DR N-PSDB; V70150.
PT Use of a tumour rejection antigen precursor - as a marker for
PT diagnosing a disorder characterised by expression of a tumour
PT rejection antigen precursor which is not tyrosinase
PS Claim 1; Column 7-9; 11pp; English.
CC A method has been developed for the diagnosis of a disorder which is
CC characterised by the expression of a tumour rejection antigen precursor
CC (TRAP) which is not tyrosinase, and which is processed to a TRA which
CC forms a complex with an HLA-A2 molecule. The present sequence represents
CC the TRAP for use in the present invention. The method comprises
CC contacting a sample from a subject with an agent specific for the
CC complex and determining the interaction between the complex and the
CC agent as a determination of the disorder. TRAP can be used for the
CC diagnosis and treatment of disorders characterised by the expression
CC of the TRAP molecules such as cancers, particularly melanoma.
SQ Sequence 118 AA;

Query Match 100.0%; Score 889; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.61e-82;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 21:48:00 2000; MasPar time 14.76 Seconds
554.356 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-2
Description: (1-118) from US09267439.pep
Perfect Score: 889
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 40.160; Variance 63.201; scale 0.635

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	102	11.5	418	11	Q9WUU6	HEPATOCYTE NUCLEAR FAC	2.04e-03
2	99	11.1	344	11	Q61394	CD2 ANTIGEN PRECURSOR	6.72e-03
3	99	11.1	623	5	O61391	GLY5A (EC 2.4.1.41) (P	6.72e-03
4	99	11.1	624	5	O61393	GLY5C (EC 2.4.1.41) (P	6.72e-03
5	99	11.1	626	5	O61392	GLY5B (EC 2.4.1.41) (P	6.72e-03
6	99	11.1	1069	4	O60245	PCDH7 (BH-PCDH)A.	6.72e-03
7	99	11.1	1072	4	O60246	PCDH7 (BH-PCDH)B.	6.72e-03
8	99	11.1	1200	4	O60247	PCDH7 (BH-PCDH)C.	6.72e-03
9	98	11.0	1035	13	O57537	NF-PROTOCADHERIN.	9.96e-03
10	97	10.9	617	10	P93050	ATPK2324.	1.47e-02
11	96	10.8	304	5	O96291	RIFIN.	2.17e-02
12	96	10.8	604	10	O04098	RECEPTOR-KINASE ISOLOG	2.17e-02
13	96	10.8	1069	11	O88185	BH-PROTOCADHERIN-A.	2.17e-02
14	95	10.7	614	10	O04557	T7N9.10.	3.19e-02
15	94	10.6	1510	5	O61802	H11E01.3 PROTEIN.	4.67e-02
16	90	10.1	107	14	Q86198	NONSTRUCTURAL PROTEIN.	2.10e-01
17	90	10.1	356	5	O18082	C06C6.1 PROTEIN.	2.10e-01
18	89	10.0	294	2	Q44382	INNER MEMBRANE PROTEIN	3.04e-01
19	89	10.0	486	5	Q9XYR0	TNF-RECEPTOR-ASSOCIATE	3.04e-01
20	89	10.0	712	4	Q9Y332	NG22.	3.04e-01
21	88	9.9	53	2	O87764	LACTOCOCCUS LACTIS CRE	4.39e-01
22	88	9.9	91	2	Q9Z911	HYPOTHETICAL 10.2 KD P	4.39e-01
23	88	9.9	215	2	Q9ZLR6	PUTATIVE.	4.39e-01
24	87	9.8	434	2	O87996	PUTATIVE INTEGRAL MEMB	6.31e-01
25	87	9.8	442	1	O50088	442AA LONG HYPOTHETICA	6.31e-01
26	87	9.8	509	5	Q9Y0H8	SERINE/THREONINE PROTE	6.31e-01
27	87	9.8	522	5	Q9XTK9	SERINE/THREONINE PROTE	6.31e-01
28	87	9.8	522	5	Q9XTM5	SERINE/THREONINE PROTE	6.31e-01
29	87	9.8	682	10	O65480	SERINE/THREONINE KINAS	6.31e-01
30	86	9.7	1829	4	O15015	KIAA0296.	9.04e-01
31	85	9.6	326	2	O68878	ABC-TYPE PERMEASE.	1.29e+00
32	84	9.4	171	4	Q13733	NA,K-ATPASE ALPHA SUBU	1.84e+00
33	84	9.4	289	14	Q9WHB9	K1 GLYCOPROTEIN.	1.84e+00
34	84	9.4	289	14	Q9WHC0	K1 GLYCOPROTEIN.	1.84e+00
35	84	9.4	306	4	Q13962	CD1A ANTIGEN (FRAGMENT	1.84e+00
36	84	9.4	334	2	Q9ZD72	HYPOTHETICAL 39.1 KD P	1.84e+00
37	84	9.4	509	5	Q94888	SERINE/THREONINE PROTE	1.84e+00
38	84	9.4	512	2	Q9ZJV5	NADH OXIDOREDUCTASE I.	1.84e+00
39	84	9.4	869	5	Q9Y0Z8	BCDNA.GH12504.	1.84e+00
40	84	9.4	1265	2	P72316	CARBON MONOXIDE-INDUCE	1.84e+00
41	83	9.3	212	11	Q61332	MB7-2 PRECURSOR.	2.62e+00
42	83	9.3	263	5	Q18346	SIMILAR TO 4-HYDROXYPH	2.62e+00
43	83	9.3	285	14	Q9WHH9	TRANSFORMING MEMBRANE	2.62e+00
44	83	9.3	285	14	Q9WHC3	K1 GLYCOPROTEIN.	2.62e+00
45	83	9.3	504	13	O13102	ACTIVIN TYPE IIB RECEP	2.62e+00

ALIGNMENTS

RESULT 1
ID Q9WUU6 PRELIMINARY; PRT; 418 AA.
AC Q9WUU6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE HEPATOCYTE NUCLEAR FACTOR 4 GAMMA.
GN HNF4G.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA TARAVIRAS S., MANTAMADIOTIS T., DONG-SI T., MINCHEVA A., LICHTER P.,
RA DREWES T., RYFFEL G.U., MONAGHAN A.P., SCHUTZ G.;
RT "Cloning and characterisation of the murine hepatocyte nuclear factor
RT 4gamma (mHNF4g) gene.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AJ242626; CAB43724.1; -.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
SQ SEQUENCE 418 AA; 46903 MW; 240062A0 CRC32;

Query Match 11.5%; Score 102; DB 11; Length 418;
Best Local Similarity 28.1%; Pred. No. 2.04e-03;
Matches 18; Conservative 20; Mismatches 23; Indels 3; Gaps 3;

```
Db     74 CRYCRLRKCFRAGMKKEAVQNERDRISTRRSTYEGSNIPSINTLAQAEVRSCQISVPSPS 133
          |:||| |: :|| | | ::::   :   |||   || :    : ::  | ::|:  ||:::
Qy     45 CWYCRRRNGYRALMDK-SLHVGTQCALTRRCPQEGFDHRD-SKVSLQE-KNCEPVVPNAP 101

Db    134 SSTD 137
          :: :
Qy    102 PAYE 105
```

RESULT 2
ID Q61394 PRELIMINARY; PRT; 344 AA.
AC Q61394;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)

```
DE   CD2 ANTIGEN PRECURSOR (T LYMPHOCYTE-LIKE ANTIGEN CD2).
GN   CD2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88140313.
RA   YAGITA H., OKUMURA K., NAKAUCHI H.;
RT   "Molecular cloning of the murine homologue of CD2. Homology of the
RT   molecule to its human counterpart T11.";
RL   J. Immunol. 140:1321-1326(1988).
DR   EMBL; M18934; AAA37397.1; -.
DR   HSSP; P08921; 1A7B.
DR   MGD; MGI:88320; Cd2.
KW   Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    344       POTENTIAL.
SQ   SEQUENCE   344 AA;  38337 MW;  AFEA9175 CRC32;

  Query Match           11.1%;  Score 99;  DB 11;  Length 344;
  Best Local Similarity  40.4%;  Pred. No. 6.72e-03;
  Matches   21;  Conservative    9;  Mismatches  18;  Indels   4;  Gaps   4;

Db    198 PEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEELEI 245
          | || :  | |   :|| |:| |:| | |:| |  |:||:  |  :|  | :
Qy     13 PKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSLHV 64


RESULT   3
ID   O61391      PRELIMINARY;      PRT;   623 AA.
AC   O61391;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GLY5A (EC 2.4.1.41) (POLYPEPTIDE N-ACETYLGALACTOSAMINYLTRANSFERASE)
DE   (PROTEIN-UDP ACETYLGALACTOSAMINYLTRANSFERASE).
GN   GLY-5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE; 98192620.
RA   HAGEN F.K., NEHRKE K.;
RT   "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT   galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence
RT   homologs from Caenorhabditis elegans.";
RL   J. Biol. Chem. 273:8268-8277(1998).
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =
CC       UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
CC   -!- COFACTOR: MANGANESE; CALCIUM.
DR   EMBL; AF031835; AAC13671.1; -.
DR   PFAM; PF00535; Glycos_transf_2; 1.
DR   PFAM; PF00652; Ricin_B_lectin; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   623 AA;  71014 MW;  C061803B CRC32;

  Query Match           11.1%;  Score 99;  DB 5;  Length 623;
  Best Local Similarity  24.1%;  Pred. No. 6.72e-03;
  Matches   19;  Conservative   20;  Mismatches  38;  Indels   2;  Gaps   2;

Db      9 ILKVLLLVPVFWICSLIF-FAATSNDSSQIGSNNDLANKIAEANFHPKAAKQDVIQGFGP 67
          || ||||:   | |     : |  : | ::|:   |: : ::  |  : :|  : :
Qy     36 ILGVLLLIGC-WYCRRRNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCE 94

Db     68 PIEPEPVVENNKVEEEEQP 86
          |: |::    :|:  |: |
Qy     95 PVVPNAPPAYEKLSAEQSP 113


RESULT   4
ID   O61393      PRELIMINARY;      PRT;   624 AA.
AC   O61393;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GLY5C (EC 2.4.1.41) (POLYPEPTIDE N-ACETYLGALACTOSAMINYLTRANSFERASE)
DE   (PROTEIN-UDP ACETYLGALACTOSAMINYLTRANSFERASE).
GN   GLY-5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE; 98192620.
RA   HAGEN F.K., NEHRKE K.;
RT   "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT   galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence
RT   homologs from Caenorhabditis elegans.";
RL   J. Biol. Chem. 273:8268-8277(1998).
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =
CC       UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
CC   -!- COFACTOR: MANGANESE; CALCIUM.
DR   EMBL; AF031837; AAC13673.1; -.
DR   PFAM; PF00535; Glycos_transf_2; 1.
DR   PFAM; PF00652; Ricin_B_lectin; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   624 AA;  71103 MW;  E4F97590 CRC32;

  Query Match           11.1%;  Score 99;  DB 5;  Length 624;
  Best Local Similarity  24.1%;  Pred. No. 6.72e-03;
  Matches   19;  Conservative   20;  Mismatches  38;  Indels   2;  Gaps   2;

Db      9 ILKVLLLVPVFWICSLIF-FAATSNDSSQIGSNNDLANKIAEANFHPKAAKQDVIQGFGP 67
          || ||||:   | |     : |  : | ::|:   |: : ::  |  : :|  : :
Qy     36 ILGVLLLIGC-WYCRRRNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCE 94

Db     68 PIEPEPVVENNKVEEEEQP 86
          |: |::    :|:  |: |
Qy     95 PVVPNAPPAYEKLSAEQSP 113


RESULT   5
ID   O61392      PRELIMINARY;      PRT;   626 AA.
AC   O61392;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   GLY5B (EC 2.4.1.41) (POLYPEPTIDE N-ACETYLGALACTOSAMINYLTRANSFERASE)
DE   (PROTEIN-UDP ACETYLGALACTOSAMINYLTRANSFERASE).
GN   GLY-5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=N2;
RX   MEDLINE; 98192620.
RA   HAGEN F.K., NEHRKE K.;
RT   "cDNA cloning and expression of a family of UDP-N-acetyl-D-
RT   galactosamine:polypeptide N-acetylgalactosaminyltransferase sequence
RT   homologs from Caenorhabditis elegans.";
RL   J. Biol. Chem. 273:8268-8277(1998).
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =
CC       UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
CC   -!- COFACTOR: MANGANESE; CALCIUM.
DR   EMBL; AF031836; AAC13672.1; -.
DR   PFAM; PF00535; Glycos_transf_2; 1.
DR   PFAM; PF00652; Ricin_B_lectin; 1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   626 AA;  71383 MW;  BE318A9D CRC32;

  Query Match           11.1%;  Score 99;  DB 5;  Length 626;
```

```
  Best Local Similarity  24.1%;  Pred. No. 6.72e-03;
  Matches    19;  Conservative   20;  Mismatches  38;  Indels   2;  Gaps   2;

Db      9 ILKVLLLVPVFWICSLIF-FAATSNDSSQIGSNNDLANKIAEANFHPKAAKQDVIQGFGP 67
          || ||||:   | |     : |  : | ::|:   |: : ::  |  : :|  : :
Qy     36 ILGVLLLIGC-WYCRRRNGYRALMDKSLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCE 94

Db     68 PIEPEPVVENNKVEEEEQP 86
          |: |::    :|:  |: |
Qy     95 PVVPNAPPAYEKLSAEQSP 113


RESULT    6
ID   O60245      PRELIMINARY;      PRT;  1069 AA.
AC   O60245;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PCDH7 (BH-PCDH)A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   Genomics 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006755; BAA25194.1; -.
DR   HSSP; P15116; 1NCJ.
DR   PROSITE; PS00232; CADHERIN; 6.
DR   PFAM; PF00028; cadherin; 6.
DR   PRINTS; PR00205; CADHERIN.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1069 AA;  116104 MW;  F1732B30 CRC32;

  Query Match           11.1%;  Score 99;  DB 4;  Length 1069;
  Best Local Similarity  50.0%;  Pred. No. 6.72e-03;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    887 GIMTVILIILIVVMARYCRSKNKNGYEA 914
          ||:|||| :|:::  :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    7
ID   O60246      PRELIMINARY;      PRT;  1072 AA.
AC   O60246;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PCDH7 (BH-PCDH)B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   Genomics 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006756; BAA25195.1; -.
DR   HSSP; P15116; 1NCJ.
DR   PROSITE; PS00232; CADHERIN; 6.
DR   PFAM; PF00028; cadherin; 6.
DR   PRINTS; PR00205; CADHERIN.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1072 AA;  116463 MW;  A3DF367C CRC32;

  Query Match           11.1%;  Score 99;  DB 4;  Length 1072;
  Best Local Similarity  50.0%;  Pred. No. 6.72e-03;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    887 GIMTVILIILIVVMARYCRSKNKNGYEA 914
          ||:|||| :|:::  :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    8
ID   O60247      PRELIMINARY;      PRT;  1200 AA.
AC   O60247;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PCDH7 (BH-PCDH)C.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RL   Genomics 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006757; BAA25196.1; -.
DR   HSSP; P15116; 1NCJ.
DR   PROSITE; PS00232; CADHERIN; 5.
DR   PFAM; PF00028; cadherin; 5.
DR   PRINTS; PR00205; CADHERIN.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1200 AA;  130337 MW;  56F1CD33 CRC32;

  Query Match           11.1%;  Score 99;  DB 4;  Length 1200;
  Best Local Similarity  50.0%;  Pred. No. 6.72e-03;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    840 GIMTVILIILIVVMARYCRSKNKNGYEA 867
          ||:|||| :|:::  :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    9
ID   O57537      PRELIMINARY;      PRT;  1035 AA.
AC   O57537;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NF-PROTOCADHERIN.
GN   NFPC.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC   Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
OC   Xenopus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   BRADLEY R.S., ESPESETH A., KINTNER C.;
RL   Curr. Biol. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AF043643; AAC41270.1; -.
DR   HSSP; P15116; 1NCJ.
DR   PROSITE; PS00232; CADHERIN; 6.
DR   PFAM; PF00028; cadherin; 6.
DR   PRINTS; PR00205; CADHERIN.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ   SEQUENCE   1035 AA;  113713 MW;  7E4D3C4E CRC32;

  Query Match           11.0%;  Score 98;  DB 13;  Length 1035;
  Best Local Similarity  50.0%;  Pred. No. 9.96e-03;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    859 GIMTVILIILVVVMARYCRAKSKNGYEA 886
          ||:|||| :|:::  :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT   10
ID   P93050      PRELIMINARY;      PRT;  617 AA.
AC   P93050;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
```

```
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   ATPK2324.
GN   RKF3 OR T9J23.16.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA C24;
RA   KERTBUNDIT S., LINACERO R., ROUZE P., GALIS I., MACAS J., DEBOECK F.,
RA   HERNALSTEENS J., DE GREVE H.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   TAKAHASHI T., MU J.-H., GASCH A., CHUA N.-H.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
RA   BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
RA   CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
RA   FRASER C.M., VENTER J.C.;
RT   "Arabidopsis thaliana chromosome II BAC T9J23 genomic sequence.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z84202; CAB06335.1; -.
DR   EMBL; AF024650; AAC50045.1; -.
DR   EMBL; AC006072; AAD13705.1; -.
DR   MENDEL; 13857; Arath;1197;13857.
DR   MENDEL; 35596; Arath;1197;35596.
DR   PFAM; PF00069; pkinase; 1.
KW   Kinase; Serine/threonine-protein kinase.
SQ   SEQUENCE   617 AA;  67223 MW;  7988346C CRC32;

  Query Match           10.9%;  Score 97;  DB 10;  Length 617;
  Best Local Similarity  37.0%;  Pred. No. 1.47e-02;
  Matches    17;  Conservative   14;  Mismatches  10;  Indels   5;  Gaps   4;

Db    216 SSFSVLLVA-SVLVITAWFWYCRRKKS-KLLKPRDTSLEAGTQSRL 259
          :::::| |  :||:: :  |||||::: : |:  | ||  |||  |
Qy     28 AGIGILTVILGVLLLIG-CWYCRRRNGYRALM--DKSLHVGTQCAL 70


RESULT   11
ID   O96291      PRELIMINARY;      PRT;   304 AA.
AC   O96291;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   RIFIN.
GN   PFB1020W.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99021743.
RA   GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAVIND L.,
RA   KOONIN E.V., SHALLOM S., MASON T., YU K., FUJII C., PEDERSON J.,
RA   SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., PERTEA M.,
RA   SALZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,
RA   FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
DR   EMBL; AE001433; AAC71991.1; -.
SQ   SEQUENCE   304 AA;  34000 MW;  6C4286F5 CRC32;

  Query Match           10.8%;  Score 96;  DB 5;  Length 304;
  Best Local Similarity  45.5%;  Pred. No. 2.17e-02;
  Matches    15;  Conservative    9;  Mismatches   6;  Indels   3;  Gaps   3;

Db    263 EPCGIAALVLLILAVVLIILYIWLYRRRKNSYK 295
          |: ||: | : ||:|:|:|   | | ||:|:|:
Qy     26 EAAGIGILTV-ILGVLLLIGC-W-YCRRRNGYR 55


RESULT   12
ID   O04098      PRELIMINARY;      PRT;   604 AA.
AC   O04098;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   RECEPTOR-KINASE ISOLOG (FRAGMENT).
GN   T20D16.21.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., PHILLIPS C.A., BRANDON R.C.,
RA   FUHRMANN J.L., WHITE O., KERLAVAGE A.R., ADAMS M.D., SOMERVILLE C.R.,
RA   VENTER J.C.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U95973; AAB65490.1; -.
DR   MENDEL; 13839; Arath;1197;13839.
DR   PFAM; PF00560; LRR; 4.
DR   PFAM; PF00069; pkinase; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   604 AA;  66816 MW;  64A9194F CRC32;

  Query Match           10.8%;  Score 96;  DB 10;  Length 604;
  Best Local Similarity  51.7%;  Pred. No. 2.17e-02;
  Matches    15;  Conservative    4;  Mismatches   9;  Indels   1;  Gaps   1;

Db    202 AGSVAG-GVLVLILLLTLLIVCWRRKRRN 229
          |   || |:| :|| : ||| ||  :|||
Qy     24 AEEAAGIGILTVILGVLLLIGCWYCRRRN 52


RESULT   13
ID   O88185      PRELIMINARY;      PRT;  1069 AA.
AC   O88185;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   BH-PROTOCADHERIN-A.
GN   PCDH7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 98277460.
RA   YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;
RT   "Cloning, expression analysis, and chromosomal localization of BH-
RT   protocadherin (PCDH7), a novel member of the cadherin superfamily.";
RL   Genomics 49:458-461(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   YOSHIDA K., HIDA M., WATANABE M., YAMAGUCHI R., TATEYAMA S.,
RA   SUGANO S.;
RT   "cDNA cloning and chromosomal mapping of mouse BH-protocadherin.";
RL   DNA Seq. 0:0-0(1998).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR   EMBL; AB006758; BAA32597.1; -.
DR   HSSP; P15116; 1NCJ.
DR   PROSITE; PS00232; CADHERIN; 5.
DR   PFAM; PF00028; cadherin; 6.
DR   PRINTS; PR00205; CADHERIN.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 21:58:39 2000; MasPar time 3.11 Seconds
68.454 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-4
Description: (1-9) from US09267439.pep
Perfect Score: 56
Sequence: 1 AAGIGILTV 9

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 15.257; Variance 45.666; scale 0.334

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	56	100.0	9	1	Y01751	Exemplary antigenic pe	1.03e+01
2	56	100.0	9	1	Y00713	Tumour antigen booster	1.03e+01
3	56	100.0	9	1	Y10567	HLA Class I motif pept	1.03e+01
4	56	100.0	9	1	Y10444	HLA Class I motif pept	1.03e+01
5	56	100.0	9	1	Y10601	HLA Class I motif pept	1.03e+01
6	56	100.0	9	1	W98938	Human leukocyte antige	1.03e+01
7	56	100.0	9	1	W39430	Human immunogenic T ce	1.03e+01
8	56	100.0	9	1	W07379	MART-1 epitope recogni	1.03e+01
9	56	100.0	9	1	W42523	Melan A/MART epitope (	1.03e+01
10	56	100.0	9	1	R84196	MART-1 melanoma antige	1.03e+01
11	56	100.0	9	1	W35512	MART-1/Melan-A protein	1.03e+01
12	56	100.0	9	1	W54602	Peptide 1 from Melan-A	1.03e+01
13	56	100.0	9	1	W77123	MART-1/MelanA syntheti	1.03e+01
14	56	100.0	9	1	W68380	Human MART1/MELAN-A pe	1.03e+01
15	56	100.0	10	1	Y01750	Exemplary antigenic pe	1.03e+01
16	56	100.0	10	1	W98934	Human leukocyte antige	1.03e+01
17	56	100.0	10	1	W98939	Human leukocyte antige	1.03e+01
18	56	100.0	10	1	W98923	Human leukocyte antige	1.03e+01
19	56	100.0	10	1	Y00712	Tumour antigen booster	1.03e+01
20	56	100.0	10	1	W98924	Human leukocyte antige	1.03e+01
21	56	100.0	10	1	W98922	Human leukocyte antige	1.03e+01
22	56	100.0	10	1	W54809	Peptide 1 from Mart-1/	1.03e+01
23	56	100.0	10	1	W22039	Antigenic MART-1 pepti	1.03e+01
24	56	100.0	10	1	W32269	Tumour rejection antig	1.03e+01
25	56	100.0	10	1	W39447	Human HLA-A*0201 immun	1.03e+01
26	56	100.0	10	1	R84198	MART-1 melanoma antige	1.03e+01
27	56	100.0	10	1	W07380	MART-1 epitope recogni	1.03e+01
28	56	100.0	10	1	W07381	MART-1 epitope recogni	1.03e+01
29	56	100.0	10	1	R84197	MART-1 melanoma antige	1.03e+01
30	56	100.0	12	1	W22038	Antigenic MART-1 pepti	1.03e+01
31	56	100.0	21	1	W00903	Human melanoma MART-1/	1.03e+01
32	56	100.0	118	1	R84212	MART-1 melanoma antige	1.03e+01
33	56	100.0	118	1	W83134	Human tumour rejection	1.03e+01
34	56	100.0	118	1	R63158	Tumour rejection antig	1.03e+01
35	53	94.6	9	1	W42524	Melan A/MART (residues	2.28e+01
36	53	94.6	9	1	W42525	Melan A/MART epitope (	2.28e+01
37	53	94.6	9	1	W42531	Melan A/MART epitope (	2.28e+01
38	53	94.6	10	1	W98936	Human leukocyte antige	2.28e+01
39	52	92.9	9	1	R84786	Modified MART-1 melano	2.96e+01
40	52	92.9	9	1	W45778	Melan A/MART epitope (	2.96e+01
41	52	92.9	9	1	W98933	Human leukocyte antige	2.96e+01
42	52	92.9	9	1	R84788	Modified MART-1 melano	2.96e+01
43	52	92.9	10	1	W98926	Human leukocyte antige	2.96e+01
44	52	92.9	10	1	W98937	Human leukocyte antige	2.96e+01
45	52	92.9	10	1	W98925	Human leukocyte antige	2.96e+01

ALIGNMENTS

RESULT 1
ID Y01751 standard; Peptide; 9 AA.
AC Y01751;
DT 25-JUN-1999 (first entry)
DE Exemplary antigenic peptide derived from Melan-A(MART-1).
KW MAGE-3; tumour associated gene; human leucocyte antigen Class II;
KW autologous CD4+ cell; MAGE-3 related disease; cancer; melanoma;
KW osteosarcoma; leukemia; carcinoma.
OS Homo sapiens.
PN WO9914326-A1.
PD 25-MAR-1999.
PF 04-SEP-1998; U18601.
PR 12-SEP-1997; US-928615.
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYVR-) UNIV VRIJE BRUSSEL.
PI Boon-Falleur T, Chaux P, Corthals J, Heirman C,
PI Luiten R, Stroobant V, Thielemans K, Van Der Bruggen P;
DR WPI; 99-244031/20.
PT Isolated peptides that bind to human leucocyte antigen class II
PT molecules
PS Disclosure; Page 29; 88pp; English.
CC The present sequence represents an exemplary tumour associated peptide
CC antigen. The specification describes a MAGE-3 tumour associated gene.
CC Peptides (Y01721-25) that bind human leucocyte antigen (HLA) Class II
CC molecules can be derived from the MAGE-3 protein. These peptides and
CC autologous CD4+ cells that bind to a complex of MAGE-3 peptide
CC and HLA Class II, are used to treat MAGE-3 related diseases,
CC particularly cancers (e.g. melanoma, osteosarcoma, leukemia and
CC various forms of carcinoma). The peptides are also used to produce
CC specific antibodies. Detection of of the peptides, e.g. in binding
CC assays, particularly with antibodies, is used for diagnosis of such
CC diseases.
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 2
ID Y00713 standard; peptide; 9 AA.
AC Y00713;
DT 12-MAY-1999 (first entry)

DE Tumour antigen booster peptide Melan-AMART-1 HLA-A2 #2.
KW Tumour antigen; booster peptide; immune response modulation; allergy;
KW immune response enhancer; tumour cell; tumour rejection antigen;
KW leukocyte antigen-presenting molecule; autoimmune disease;
KW allograft rejection.
OS Homo sapiens.
PN WO9858956-A2.
PD 30-DEC-1998.
PF 19-JUN-1998; U12894.
PR 23-JUN-1997; US-880979.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, Uyttenhove C, Warnier G;
DR WPI; 99-105612/09.
PT Immunization methods using viruses expressing antigen for priming
PT and booster immunizations - useful for modulating immune responses
PT against antigen, e.g. enhancing immune response against tumour cells
PT expressing tumour rejection antigens
PS Disclosure; Page 10; 33pp; English.
CC This sequence represents a tumour antigen booster peptide that can be
CC used in the method of the invention. The method is for for modulating an
CC immune response in a mammal against an antigen, and comprises:
CC (A) inducing an immune response by: (i) administering a virus containing
CC a nucleic acid molecule encoding the antigen or its precursor to generate
CC an immune response; and (ii) administering at least one booster dose
CC comprising a peptide including the antigen, in an adjuvant, in a combined
CC amount effective to enhance the initial immune response; or
CC (B) reducing an immune response as defined for (A) but using a
CC non-adjuvant with the peptide which includes the antigen, in an amount
CC effective to reduce the initial immune response. Method (A) is used to
CC enhance the immune response against tumour cells expressing tumour
CC rejection antigens, and against pathogens in subjects having human
CC leukocyte antigen-presenting molecules. Method (B) is used to reduce the
CC immune response in allergy, autoimmune disease, and allograft rejection.
CC Method (A) provides an immunisation method which, unlike prior art, is
CC not limited by the host immune response against viral vectors.
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 3
ID Y10567 standard; Peptide; 9 AA.
AC Y10567;
DT 12-MAY-1999 (first entry)
DE HLA Class I motif peptide SEQ ID NO:497.
KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW immunisation; tumour; infectious disease; immunotherapy; cancer;
KW malignant melanoma; viral disease; hepatitis; AIDS.
OS Synthetic.
OS Homo sapiens.
PN WO9902183-A2.
PD 21-JAN-1999.
PF 10-JUL-1998; U14289.
PR 10-DEC-1997; US-988320.
PR 10-JUL-1997; CA-209815.
PA (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI Kuendig TM, Simard JJL;
DR WPI; 99-120514/10.
PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT of antigen in the lymphatic system of a mammal so as to provide a
PT sustained CTL response, used to treat, e.g. AIDS
PS Disclosure; Page 47; 199pp; English.
CC The present invention describes a method of inducing and/or sustaining
CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC method comprises: (a) delivering an antigen to the mammal at a level to
CC induce an immunological CTL response in the mammal; and (b) maintaining
CC the level of the antigen in the mammal's lymphatic system to maintain
CC the immunologic CTL response. The method can be used for the delivery of
CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC gene antigen, or a viral antigen. They can be used for the treatment of
CC disease such as cancer, e.g. malignant melanoma or infectious disease,
CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC to the lymphatic system provides for potent CTL stimulation that takes
CC place in the milieu of the lymphoid organ, and it sustains stimulation
CC that is necessary to keep CTL active, cytotoxic and recirculating
CC through the body. Y10071 to Y10639 represent examples of peptide
CC antigens given in the present invention.
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 4
ID Y10444 standard; Peptide; 9 AA.
AC Y10444;
DT 12-MAY-1999 (first entry)
DE HLA Class I motif peptide SEQ ID NO:374.
KW Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW immunisation; tumour; infectious disease; immunotherapy; cancer;
KW malignant melanoma; viral disease; hepatitis; AIDS.
OS Synthetic.
OS Homo sapiens.
PN WO9902183-A2.
PD 21-JAN-1999.
PF 10-JUL-1998; U14289.
PR 10-DEC-1997; US-988320.
PR 10-JUL-1997; CA-209815.
PA (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI Kuendig TM, Simard JJL;
DR WPI; 99-120514/10.
PT Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT of antigen in the lymphatic system of a mammal so as to provide a
PT sustained CTL response, used to treat, e.g. AIDS
PS Disclosure; Page 40; 199pp; English.
CC The present invention describes a method of inducing and/or sustaining
CC an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC method comprises: (a) delivering an antigen to the mammal at a level to
CC induce an immunological CTL response in the mammal; and (b) maintaining
CC the level of the antigen in the mammal's lymphatic system to maintain
CC the immunologic CTL response. The method can be used for the delivery of
CC e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC gene antigen, or a viral antigen. They can be used for the treatment of
CC disease such as cancer, e.g. malignant melanoma or infectious disease,
CC e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC to the lymphatic system provides for potent CTL stimulation that takes
CC place in the milieu of the lymphoid organ, and it sustains stimulation
CC that is necessary to keep CTL active, cytotoxic and recirculating
CC through the body. Y10071 to Y10639 represent examples of peptide
CC antigens given in the present invention.
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 5
ID Y10601 standard; Peptide; 9 AA.

```
AC   Y10601;
DT   12-MAY-1999  (first entry)
DE   HLA Class I motif peptide SEQ ID NO:531.
KW   Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW   immunisation; tumour; infectious disease; immunotherapy; cancer;
KW   malignant melanoma; viral disease; hepatitis; AIDS.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9902183-A2.
PD   21-JAN-1999.
PF   10-JUL-1998; U14289.
PR   10-DEC-1997; US-988320.
PR   10-JUL-1997; CA-209815.
PA   (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI   Kuendig TM, Simard JJL;
DR   WPI; 99-120514/10.
PT   Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT   of antigen in the lymphatic system of a mammal so as to provide a
PT   sustained CTL response, used to treat, e.g. AIDS
PS   Disclosure; Page 49; 199pp; English.
CC   The present invention describes a method of inducing and/or sustaining
CC   an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC   method comprises: (a) delivering an antigen to the mammal at a level to
CC   induce an immunological CTL response in the mammal; and (b) maintaining
CC   the level of the antigen in the mammal's lymphatic system to maintain
CC   the immunologic CTL response. The method can be used for the delivery of
CC   e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC   an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC   gene antigen, or a viral antigen. They can be used for the treatment of
CC   disease such as cancer, e.g. malignant melanoma or infectious disease,
CC   e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC   to the lymphatic system provides for potent CTL stimulation that takes
CC   place in the milieu of the lymphoid organ, and it sustains stimulation
CC   that is necessary to keep CTL active, cytotoxic and recirculating
CC   through the body. Y10071 to Y10639 represent examples of peptide
CC   antigens given in the present invention.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT    6
ID   W98938 standard; peptide; 9 AA.
AC   W98938;
DT   06-MAY-1999  (first entry)
DE   Human leukocyte antigen A2 molecule binding peptide SEQ ID NO:2.
KW   Human leukocyte antigen; HLA; HLA-A2 binding peptide; T cell;
KW   cytolytic T cell; CTL.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9858951-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; U12879.
PR   16-APR-1998; US-061388.
PR   23-JUN-1997; US-880963.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Cerottini J, Romero P, Valmori D;
DR   WPI; 99-105609/09.
PT   New decamer peptides which bind to HLA molecules - useful to
PT   identify HLA-A2 positive cells and provoke T cells
PS   Claim 17; Page 9; 45pp; English.
CC   The present invention describes peptides which bind to an HLA-A2
CC   molecule and have Val at the carboxy terminus, and either: (a) Ala, Tyr
CC   or Phe at the amino terminus, and Ala at position 2 (P1); or (b) Glu at
CC   the amino terminus, and Ala, Leu, or Met at positions 2 and 3, with the
CC   proviso that Ala is not at both positions (P2). The present sequence
CC   represents an HLA-A2 binding peptide. The peptides of the present
CC   invention are used to identify HLA-A2 positive cells, provoke T cells,
CC   and determine the presence of particular T cells including cytolytic
CC   T cells (CTLs). They provide a better target than the prior art
CC   CTL-stimulating peptide.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT    7
ID   W39430 standard; peptide; 9 AA.
AC   W39430;
DT   11-JUN-1998  (first entry)
DE   Human immunogenic T cell epitope 1.
KW   T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW   vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW   disease; anti-tumour; anti-viral.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9741440-A1.
PD   06-NOV-1997.
PF   28-APR-1997; NL0229.
PR   23-DEC-1996; EP-203670.
PR   26-APR-1996; EP-201145.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PA   (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
PI   Kast WM, Melief CJM, Offringa R, Toes REM, Van Der Burg SH;
DR   WPI; 97-549891/50.
PT   Method of selecting T cell peptide epitope(s) - by measuring the
PT   stability of HLA class I-peptide complexes on intact B cells
PS   Disclosure; Page 6; 109pp; English.
CC   Peptides W39430-W39734 are used in a novel method for the selection of
CC   immunogenic T-cell peptide epitopes present in polypeptide antigens.
CC   Peptides W39430 and W39431 are derived from MART-1. The method involves
CC   the identification of peptide sequences capable of binding to an HLA
CC   (human leukocyte antigen) class I molecule and measuring the binding of
CC   this epitope peptide to the HLA class I peptide. The stability of binding
CC   of the peptide and MHC (major histocompatablilty complex) class I
CC   molecule is measured on intact human B cells carrying the MHC molecule at
CC   their cell surfaces. The method can be used to select peptide epitopes
CC   for generating vaccines against a disease associated with the
CC   polypeptide, e.g. cancers or AIDS. The peptide epitopes are especially
CC   T-cell peptide epitopes with strong anti-tumour and anti-viral immune
CC   responses.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT    8
ID   W07379 standard; Peptide; 9 AA.
AC   W07379;
DT   28-JUL-1997  (first entry)
DE   MART-1 epitope recognised by melanoma specific T cell receptor.
KW   T cell; receptor; lymphocyte; alpha; beta chain; V; variable;
KW   J; joining; D; diversity; gene segment; probe; detection;
KW   recombination; melanoma; cancer; neoplasia; tumour; diagnosis;
KW   MART; Melanoma Antigen Recognised by T lymphocyte.
OS   Homo sapiens.
PN   WO9630516-A1.
PD   03-OCT-1996.
```

PF 27-MAR-1996; U04143.
PR 27-MAR-1995; US-411098.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Hwu P, Nishimura M, Rosenberg SA;
DR WPI; 96-485449/48.
PT T cell receptor alpha and/or beta chains, and related nucleic acids
PT useful in pharmaceutical compsns. to prevent or treat cancer,
PT partic. lung, melanoma, ovarian, colon, brain or kidney tumours
PS Example 3; Page 11; 125pp; English
CC W07378-W07381 are MART-1 epitopes, M9-1, M9-2, M10-3 and M10-4
CC respectively, that are recognised by melanoma specific T lymphocyte
CC receptors (TCRs). Melanoma-specific TCRs comprising an alpha and
CC beta chain were made. Nucleic acids from either of these chains can be
CC used as probes for the detection of expression of rearranged genes
CC encoding tumour-associated antigens. The nucleic acids may also be used
CC to create transgenic animals, useful as biological models to study cancer
CC and evaluate diagnostic and therapeutic methods for the treatment of
CC cancers, particularly melanomas. Antibodies (Abs) may be raised against
CC alpha and beta chain polypeptides and used to detect native or denatured
CC TCRs and/or alterations in expression levels of T cells carrying
CC melanoma-specific TCRs. Abs can also purify and enrich T cells carrying
CC the above receptors, which can then be administered therapeutically to
CC mammals. Anti-idiotype antibodies can be used to assess the level of a
CC specific T cell carrying these receptors in a mammal being treated using
CC these methods. Host cells and vectors carrying nucleic acid encoding
CC a TCR (or individual alpha or beta chain fragment) are useful in
CC pharmaceutical compositions to prevent or treat cancer in a mammal, e.g.
CC lung, melanoma, ovarian, colon, brain or kidney tumours.
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 9
ID W42523 standard; peptide; 9 AA.
AC W42523;
DT 22-JUN-1998 (first entry)
DE Melan A/MART epitope (residues 27-35).
KW Metastatic melanoma; peptide analogue; vaccine; cancer; diagnosis;
KW antigen; CTL; immunogenic; viral disease; gp 100; Melan A/MART-1.
OS Synthetic.
OS Homo sapiens.
PN WO9802538-A1.
PD 22-JAN-1998.
PF 08-JUL-1997; E03712.
PR 11-JUL-1996; EP-201945.
PA (ALKU) AKZO NOBEL NV.
PI Adema GJ, Figdor CG;
DR WPI; 98-110586/10.
PT Melanoma associated peptide analogues - useful in vaccines against
PT melanoma
PS Example 1; Page 28; 47pp; English.
CC This sequence is shown in the specification. The invention relates to
CC peptides, which are immunogenic with lymphocytes directed against
CC metastatic melanomas. They are characterised in that they comprise at
CC least a part of the following sequence, where the amino acid at position
CC 2 or 8 is substituted: Lys-Thr-Trp-Gly-Gln-Tyr-Trp-Gln-Val. Vaccines
CC comprising the peptide, an epitope of the peptide, nucleotide sequence
CC encoding the peptide, or an antigen presenting cell preloaded with the
CC peptide or antibody as above, are useful for cancer, particularly
CC melanoma, treatment. The peptides can also be used to generate antigen
CC reactive tumour infiltrating lymphocytes, which can also be used in
CC vaccines. The peptides can be exploited to elicit native epitope-reactive
CC CTL. Usage of the peptides with improved immunogenicity may contribute
CC to the development of CTL-epitope based vaccines in viral disease and
CC cancer.
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 10
ID R84196 standard; Peptide; 9 AA.
AC R84196;
DT 20-APR-1996 (first entry)
DE MART-1 melanoma antigen immunogenic peptide M9-2.
KW MART-1; M9-2; melanoma antigen recognised by T-cells; melanoma;
KW metastatic melanoma; tumour-associated antigen;
KW immunogenic peptide; diagnosis; prognosis; prophylaxis;
KW therapy; vaccine.
OS Synthetic.
PN WO9529193-A2.
PD 02-NOV-1995.
PF 21-APR-1995; U05063.
PR 22-APR-1994; US-231565.
PR 05-APR-1995; US-417174.
PA (USSH) US SEC DEPT HEALTH.
PI Kawakami Y, Rosenberg SA;
DR WPI; 95-382963/49.
PT DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT vectors, host cells and antibodies, used to detect, treat and
PT immunise animal against melanoma.
PS Claim 12; Page 117; 184pp; English.
CC Immunogenic peptide M9-2 is based on the melanoma antigen (MART-1)
CC (see R84212). M9-2 may be modified to improve immunogenicity
CC (see R84783-R84800) and used in medicaments for the treatment or
CC prevention (by immunization) of melanoma. Antibodies against MART-1
CC and its immunogenic peptides may be used in the detection and
CC isolation of MART-1 from a sample, the detection of which is
CC indicative of a disease state (melanoma or metastatic melanoma).
SQ Sequence 9 AA;

Query Match 100.0%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTV 9
|||||||||
Qy 1 AAGIGILTV 9

RESULT 11
ID W35512 standard; peptide; 9 AA.
AC W35512;
DT 22-APR-1998 (first entry)
DE MART-1/Melan-A protein peptide SEQ ID NO:44 from WO9738011.
KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
KW scaffold; inhibition; metastasis; wound healing; solid phase.
OS Unidentified.
PN WO9738011-A1.
PD 16-OCT-1997.
PF 03-APR-1997; D00146.
PR 03-APR-1996; DK-000398.
PA (PEPR-) PEPRESEARCH AS.
PI Heegaard PMH, Jakobsen PH;
DR WPI; 97-512645/47.
PT Non-dendritic peptide carrier linked to a solid phase - useful as a
PT diagnostic agent and as a scaffold for production of chemical
PT derivatives
PS Example 26; Page 146; 262pp; English.
CC A non-dendritic peptide carrier (A) has been developed which is coupled
CC through a linker to a solid phase, forming a complex of (A)-solid phase.
CC Where (A) comprises 10-50 amino acids capable of forming a secondary
CC structure in a benign buffer after liberation from the solid phase, and

```
CC   further the (A)-solid phase complex comprises an immunogenic substance
CC   and/or an immune mediator coupled on (A). The present sequence
CC   represents a peptide used in an example from the present invention. An
CC   (A)-solid phase complex can be used as a scaffold for the production of
CC   chemical derivatives, characterised by covalently attaching molecules at
CC   attachment points. Alternatively (A) is used as a scaffold-peptide for
CC   the incorporation into an Immunostimulating Complex (Iscom) resulting an
CC   (A)-Iscom complex which is used for the chemical coupling of antigenic
CC   substances in an aqueous solution by conjugation. (A) derivatised with
CC   one or more peptides having fibronectin-, laminin- or vitronectin-like
CC   binding activities can be used for the promotion of cell-attachment to
CC   plastic surfaces, in particular to inhibit tumour growth and metastasis,
CC   and for promotion of wound healing. Also a derivatised (A) can be used
CC   for the selection of specifically-binding aptamers or as a diagnostic
CC   agent. Such diagnostic-(A) molecules could be used to detect molecules
CC   derived from or indicative of pregnancy or of a disease, such as an
CC   infectious, autoimmune or cancerous disease.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   12
ID   W54602 standard; peptide; 9 AA.
AC   W54602;
DT   25-SEP-1998  (first entry)
DE   Peptide 1 from Melan-A/Mart-1.
KW   Mannose; antigen; antigen-presenting cell; mannosylated peptide; T cell;
KW   vaccine; treatment.
OS   Synthetic.
PN   WO9813378-A1.
PD   02-APR-1998.
PF   25-SEP-1997; NL0536.
PR   26-SEP-1996; EP-202701.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Drijfhout JW, Koning F;
DR   WPI; 98-230631/20.
PT   Increasing uptake and presentation of antigen(s) - by adding mannose
PT   residue(s) to antigen for increasing T cell response, useful in,
PT   e.g. vaccines against viral infection(s)
PS   Disclosure; Page 24; 47pp; English.
CC   The peptides W54559-W54809 are examples of peptides to which at least 1
CC   (preferably 2) mannose can be attached to increase their uptake as
CC   antigens by antigen-presenting cells.  Uptake of agonist mannosylated
CC   peptides will increase the T cell response, whereas uptake of antagonist
CC   peptides blocks the T cell response.  Blocking binding of immunogenic
CC   autoantigens can be used in treatment of type I diabetes, rheumatoid
CC   arthritis, graft rejection etc., also to induce T-cell non-
CC   responsiveness.  Vaccines containing mannosylated antigen are used to
CC   prevent or treat infections by, e.g. bacteria, viruses, fungi, helminths
CC   and parasites.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   13
ID   W77123 standard; peptide; 9 AA.
AC   W77123;
DT   16-NOV-1998  (first entry)
DE   MART-1/MelanA synthetic peptide epitope 1.
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
OS   Synthetic.
PN   WO9833810-A2.
PD   06-AUG-1998.
PF   29-JAN-1998; U01592.
PR   30-JAN-1997; US-037781.
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
PI   Engelhard VH, Hunt DF, Kittlesen D, Slingluff CL;
DR   WPI; 98-437388/37.
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
PS   Disclosure; Page 27; 93pp; English.
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   14
ID   W68380 standard; peptide; 9 AA.
AC   W68380;
DT   14-OCT-1998  (first entry)
DE   Human MART1/MELAN-A peptide binds HLA-A2.
KW   Antigen; major histocompatibility complex; MHC; lymphocyte; detection;
KW   immobilisation; cytotoxic T-cell; tumour; leukaemia; lymphoma;
KW   viral infection.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9744667-A2.
PD   27-NOV-1997.
PF   21-MAY-1997; F00892.
PR   21-MAY-1996; US-651925.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PI   Abastado J, Kourilsky P, Langlade-Demoyen P, Lone Y;
DR   WPI; 98-018653/02.
PT   Detection, purification and elimination of antigen-specific
PT   lymphocytes - for producing cytotoxic T cells for immuno-therapy of
PT   cancers and viral infection
PS   Disclosure; Page 30; 222pp; French.
CC   Peptides W68301-W68384 are examples of antigens (Ag) which can be loaded
CC   onto recombinantly produced major histocompatibility complex (MHC)
CC   molecules in a method of detecting antigen-specific lymphocytes.  The
CC   MHC-antigen complex is then immobilised on a solid support and a sample
CC   containing cells recognising the MHC-Ag complex may be isolated.  This
CC   peptide is derived from the human MART1/MELAN-A protein and binds the
CC   human leukocyte antigen A2 (HLA-A2). A similar method is used to isolate,
CC   purify or eliminate Ag-specific T-cells or to produce Ag-specific
CC   cytotoxic T-cells (CTC).  The method is also used to detect and quantify
CC   tumour-specific T-cells and to generate CTC for specific killing of
CC   tumour cells (solid tumours, leukaemia or lymphoma) by injection into
CC   a human or animal, but also for treating viral infections.
SQ   Sequence   9 AA;

  Query Match          100.0%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.03e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9
```

RESULT 15
ID Y01750 standard; Peptide; 10 AA.
AC Y01750;
DT 25-JUN-1999 (first entry)
DE Exemplary antigenic peptide derived from Melan-A(MART-1).
KW MAGE-3; tumour associated gene; human leucocyte antigen Class II;
KW autologous CD4+ cell; MAGE-3 related disease; cancer; melanoma;
KW osteosarcoma; leukemia; carcinoma.
OS Homo sapiens.
PN WO9914326-A1.
PD 25-MAR-1999.
PF 04-SEP-1998; U18601.
PR 12-SEP-1997; US-928615.
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYVR-) UNIV VRIJE BRUSSEL.
PI Boon-Falleur T, Chaux P, Corthals J, Heirman C,
PI Luiten R, Stroobant V, Thielemans K, Van Der Bruggen P;
DR WPI; 99-244031/20.
PT Isolated peptides that bind to human leucocyte antigen class II
PT molecules
PS Disclosure; Page 29; 88pp; English.
CC The present sequence represents an exemplary tumour associated peptide
CC antigen. The specification describes a MAGE-3 tumour associated gene.
CC Peptides (Y01721-25) that bind human leucocyte antigen (HLA) Class II
CC molecules can be derived from the MAGE-3 protein. These peptides and
CC autologous CD4+ cells that bind to a complex of MAGE-3 peptide
CC and HLA Class II, are used to treat MAGE-3 related diseases,
CC particularly cancers (e.g. melanoma, osteosarcoma, leukemia and
CC various forms of carcinoma). The peptides are also used to produce
CC specific antibodies. Detection of of the peptides, e.g. in binding
CC assays, particularly with antibodies, is used for diagnosis of such
CC diseases.
SQ Sequence 10 AA;

Query Match 100.0%; Score 56; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.03e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      2 AAGIGILTV 10
          |||||||||
Qy      1 AAGIGILTV 9
```

Search completed: Fri May 5 21:59:14 2000
Job time : 35 secs.

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:07:09 2000; MasPar time 3.08 Seconds
77.018 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-17
Description: (1-10) from US09267439.pep
Perfect Score: 62
Sequence: 1 EAAGIGILTV 10

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 15.635; Variance 47.333; scale 0.330

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	62	100.0	10	1	W98939	Human leukocyte antige	2.79e+00
2	62	100.0	10	1	Y01750	Exemplary antigenic pe	2.79e+00
3	62	100.0	10	1	Y00712	Tumour antigen booster	2.79e+00
4	62	100.0	10	1	W39447	Human HLA-A*0201 immun	2.79e+00
5	62	100.0	10	1	W07380	MART-1 epitope recogni	2.79e+00
6	62	100.0	10	1	W32269	Tumour rejection antig	2.79e+00
7	62	100.0	10	1	W54809	Peptide 1 from Mart-1/	2.79e+00
8	62	100.0	10	1	W22039	Antigenic MART-1 pepti	2.79e+00
9	62	100.0	10	1	R84197	MART-1 melanoma antige	2.79e+00
10	62	100.0	118	1	R84212	MART-1 melanoma antige	2.79e+00
11	62	100.0	118	1	R63158	Tumour rejection antig	2.79e+00
12	62	100.0	118	1	W83134	Human tumour rejection	2.79e+00
13	59	95.2	10	1	W98936	Human leukocyte antige	6.17e+00
14	56	90.3	9	1	Y10567	HLA Class I motif pept	1.36e+01
15	56	90.3	9	1	Y00713	Tumour antigen booster	1.36e+01
16	56	90.3	9	1	R84196	MART-1 melanoma antige	1.36e+01
17	56	90.3	9	1	W35512	MART-1/Melan-A protein	1.36e+01
18	56	90.3	9	1	W68380	Human MART1/MELAN-A pe	1.36e+01
19	56	90.3	9	1	W77123	MART-1/MelanA syntheti	1.36e+01
20	56	90.3	9	1	W54602	Peptide 1 from Melan-A	1.36e+01
21	56	90.3	9	1	W07379	MART-1 epitope recogni	1.36e+01
22	56	90.3	9	1	Y01751	Exemplary antigenic pe	1.36e+01
23	56	90.3	9	1	Y10444	HLA Class I motif pept	1.36e+01
24	56	90.3	9	1	W98938	Human leukocyte antige	1.36e+01
25	56	90.3	9	1	Y10601	HLA Class I motif pept	1.36e+01
26	56	90.3	9	1	W39430	Human immunogenic T ce	1.36e+01
27	56	90.3	9	1	W42523	Melan A/MART epitope (	1.36e+01
28	56	90.3	10	1	W98922	Human leukocyte antige	1.36e+01
29	56	90.3	10	1	W98934	Human leukocyte antige	1.36e+01
30	56	90.3	10	1	W98926	Human leukocyte antige	1.36e+01
31	56	90.3	10	1	W98928	Human leukocyte antige	1.36e+01
32	56	90.3	10	1	W98924	Human leukocyte antige	1.36e+01
33	56	90.3	10	1	W98923	Human leukocyte antige	1.36e+01
34	56	90.3	10	1	R84198	MART-1 melanoma antige	1.36e+01
35	56	90.3	10	1	W07381	MART-1 epitope recogni	1.36e+01
36	56	90.3	12	1	W22038	Antigenic MART-1 pepti	1.36e+01
37	56	90.3	21	1	W00903	Human melanoma MART-1/	1.36e+01
38	55	88.7	10	1	W98925	Human leukocyte antige	1.76e+01
39	55	88.7	10	1	W98927	Human leukocyte antige	1.76e+01
40	53	85.5	9	1	W42524	Melan A/MART (residues	2.94e+01
41	53	85.5	9	1	W42531	Melan A/MART epitope (	2.94e+01
42	53	85.5	9	1	W42525	Melan A/MART epitope (	2.94e+01
43	52	83.9	9	1	W98933	Human leukocyte antige	3.80e+01
44	52	83.9	9	1	W98932	Human leukocyte antige	3.80e+01
45	52	83.9	9	1	R84788	Modified MART-1 melano	3.80e+01

ALIGNMENTS

RESULT 1
ID W98939 standard; peptide; 10 AA.
AC W98939;
DT 06-MAY-1999 (first entry)
DE Human leukocyte antigen A2 molecule binding peptide SEQ ID NO:1.
KW Human leukocyte antigen; HLA; HLA-A2 binding peptide; T cell;
KW cytolytic T cell; CTL.
OS Synthetic.
OS Homo sapiens.
PN WO9858951-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12879.
PR 16-APR-1998; US-061388.
PR 23-JUN-1997; US-880963.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Cerottini J, Romero P, Valmori D;
DR WPI; 99-105609/09.
PT New decamer peptides which bind to HLA molecules - useful to
PT identify HLA-A2 positive cells and provoke T cells
PS Claim 18; Page 6; 45pp; English.
CC The present invention describes peptides which bind to an HLA-A2
CC molecule and have Val at the carboxy terminus, and either: (a) Ala, Tyr
CC or Phe at the amino terminus, and Ala at position 2 (P1); or (b) Glu at
CC the amino terminus, and Ala, Leu, or Met at positions 2 and 3, with the
CC proviso that Ala is not at both positions (P2). The present sequence
CC represents an HLA-A2 binding peptide. The peptides of the present
CC invention are used to identify HLA-A2 positive cells, provoke T cells,
CC and determine the presence of particular T cells including cytolytic
CC T cells (CTLs). They provide a better target than the prior art
CC CTL-stimulating peptide.
SQ Sequence 10 AA;

Query Match 100.0%; Score 62; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.79e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      1 EAAGIGILTV 10
          ||||||||||
Qy      1 EAAGIGILTV 10
```

RESULT 2
ID Y01750 standard; Peptide; 10 AA.
AC Y01750;
DT 25-JUN-1999 (first entry)
DE Exemplary antigenic peptide derived from Melan-A(MART-1).
KW MAGE-3; tumour associated gene; human leucocyte antigen Class II;

KW autologous CD4+ cell; MAGE-3 related disease; cancer; melanoma;
KW osteosarcoma; leukemia; carcinoma.
OS Homo sapiens.
PN WO9914326-A1.
PD 25-MAR-1999.
PF 04-SEP-1998; U18601.
PR 12-SEP-1997; US-928615.
PA (LUDW-) LUDWIG INST CANCER RES.
PA (UYVR-) UNIV VRIJE BRUSSEL.
PI Boon-Falleur T, Chaux P, Corthals J, Heirman C,
PI Luiten R, Stroobant V, Thielemans K, Van Der Bruggen P;
DR WPI; 99-244031/20.
PT Isolated peptides that bind to human leucocyte antigen class II
PT molecules
PS Disclosure; Page 29; 88pp; English.
CC The present sequence represents an exemplary tumour associated peptide
CC antigen. The specification describes a MAGE-3 tumour associated gene.
CC Peptides (YO1721-25) that bind human leucocyte antigen (HLA) Class II
CC molecules can be derived from the MAGE-3 protein. These peptides and
CC autologous CD4+ cells that bind to a complex of MAGE-3 peptide
CC and HLA Class II, are used to treat MAGE-3 related diseases,
CC particularly cancers (e.g. melanoma, osteosarcoma, leukemia and
CC various forms of carcinoma). The peptides are also used to produce
CC specific antibodies. Detection of of the peptides, e.g. in binding
CC assays, particularly with antibodies, is used for diagnosis of such
CC diseases.
SQ Sequence 10 AA;

Query Match 100.0%; Score 62; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.79e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 EAAGIGILTV 10
||||||||||
Qy 1 EAAGIGILTV 10

RESULT 3
ID Y00712 standard; peptide; 10 AA.
AC Y00712;
DT 12-MAY-1999 (first entry)
DE Tumour antigen booster peptide Melan-AMART-1 HLA-A2 #1.
KW Tumour antigen; booster peptide; immune response modulation; allergy;
KW immune response enhancer; tumour cell; tumour rejection antigen;
KW leukocyte antigen-presenting molecule; autoimmune disease;
KW allograft rejection.
OS Homo sapiens.
PN WO9858956-A2.
PD 30-DEC-1998.
PF 19-JUN-1998; U12894.
PR 23-JUN-1997; US-880979.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, Uyttenhove C, Warnier G;
DR WPI; 99-105612/09.
PT Immunization methods using viruses expressing antigen for priming
PT and booster immunizations - useful for modulating immune responses
PT against antigen, e.g. enhancing immune response against tumour cells
PT expressing tumour rejection antigens
PS Disclosure; Page 10; 33pp; English.
CC This sequence represents a tumour antigen booster peptide that can be
CC used in the method of the invention. The method is for for modulating an
CC immune response in a mammal against an antigen, and comprises:
CC (A) inducing an immune response by: (i) administering a virus containing
CC a nucleic acid molecule encoding the antigen or its precursor to generate
CC an immune response; and (ii) administering at least one booster dose
CC comprising a peptide including the antigen, in an adjuvant, in a combined
CC amount effective to enhance the initial immune response; or
CC (B) reducing an immune response as defined for (A) but using a
CC non-adjuvant with the peptide which includes the antigen, in an amount
CC effective to reduce the initial immune response. Method (A) is used to
CC enhance the immune response against tumour cells expressing tumour
CC rejection antigens, and against pathogens in subjects having human
CC leukocyte antigen-presenting molecules. Method (B) is used to reduce the
CC immune response in allergy, autoimmune disease, and allograft rejection.
CC Method (A) provides an immunisation method which, unlike prior art, is
CC not limited by the host immune response against viral vectors.
SQ Sequence 10 AA;

Query Match 100.0%; Score 62; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.79e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 EAAGIGILTV 10
||||||||||
Qy 1 EAAGIGILTV 10

RESULT 4
ID W39447 standard; peptide; 10 AA.
AC W39447;
DT 11-JUN-1998 (first entry)
DE Human HLA-A*0201 immunogenic peptide 10-mer.
KW T cell epitope; immune response; human leukocyte antigen; HLA Class I;
KW vaccine; immunogenic; major histocompatablilty complex; MHC; B cell;
KW disease; anti-tumour; anti-viral.
OS Synthetic.
OS Homo sapiens.
PN WO9741440-A1.
PD 06-NOV-1997.
PF 28-APR-1997; NL0229.
PR 23-DEC-1996; EP-203670.
PR 26-APR-1996; EP-201145.
PA (UYLE-) RIJKSUNIV LEIDEN.
PA (SCIS-) SCI SEED CAPITAL INVESTMENTS BV.
PI Kast WM, Melief CJM, Offringa R, Toes REM, Van Der Burg SH;
DR WPI; 97-549891/50.
PT Method of selecting T cell peptide epitope(s) - by measuring the
PT stability of HLA class I-peptide complexes on intact B cells
PS Example 3; Page 29; 109pp; English.
CC Peptides W39430-W39734 are used in a novel method for the selection of
CC immunogenic T-cell peptide epitopes present in polypeptide antigens. The
CC method involves the identification of peptide sequences capable of
CC binding to an HLA (human leukocyte antigen) class I molecule and
CC measuring the binding of this epitope peptide to the HLA class I peptide.
CC The stability of binding of the peptide and MHC (major histocompatablilty
CC complex) class I molecule is measured on intact human B cells carrying
CC the MHC molecule at their cell surfaces. The method can be used to select
CC peptide epitopes for generating vaccines against a disease associated
CC with the polypeptide, e.g. cancers or AIDS. The peptide epitopes are
CC especially T-cell peptide epitopes with strong anti-tumour and anti-viral
CC immune responses. Peptide W39447 is an immunodominant peptide-epitope
CC presented by HLA-A*0201-positive melanoma cells and displays considerable
CC binding to HLA-A*0201 in assays.
SQ Sequence 10 AA;

Query Match 100.0%; Score 62; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 2.79e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 EAAGIGILTV 10
||||||||||
Qy 1 EAAGIGILTV 10

RESULT 5
ID W07380 standard; Peptide; 10 AA.
AC W07380;
DT 28-JUL-1997 (first entry)
DE MART-1 epitope recognised by melanoma specific T cell receptor.
KW T cell; receptor; lymphocyte; alpha; beta chain; V; variable;
KW J; joining; D; diversity; gene segment; probe; detection;
KW recombination; melanoma; cancer; neoplasia; tumour; diagnosis;
KW MART; Melanoma Antigen Recognised by T lymphocyte.
OS Homo sapiens.
PN WO9630516-A1.
PD 03-OCT-1996.

```
PF   27-MAR-1996; U04143.
PR   27-MAR-1995; US-411098.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Hwu P,  Nishimura M,  Rosenberg SA;
DR   WPI; 96-485449/48.
PT   T cell receptor alpha and/or beta chains, and related nucleic acids
PT   - useful in pharmaceutical compsns. to prevent or treat cancer,
PT   partic. lung, melanoma, ovarian, colon, brain or kidney tumours
PS   Example 3; Page 11; 125pp; English
CC   W07378-W07381 are MART-1 epitopes, M9-1, M9-2, M10-3 and M10-4
CC   respectively, that are recognised by melanoma specific T lymphocyte
CC   receptors (TCRs). Melanoma-specific TCRs comprising an alpha and
CC   beta chain were made. Nucleic acids from either of these chains can be
CC   used as probes for the detection of expression of rearranged genes
CC   encoding tumour-associated antigens. The nucleic acids may also be used
CC   to create transgenic animals, useful as biological models to study cancer
CC   and evaluate diagnostic and therapeutic methods for the treatment of
CC   cancers, particularly melanomas. Antibodies (Abs) may be raised against
CC   alpha and beta chain polypeptides and used to detect native or denatured
CC   TCRs and/or alterations in expression levels of T cells carrying
CC   melanoma-specific TCRs. Abs can also purify and enrich T cells carrying
CC   the above receptors, which can then be administered therapeutically to
CC   mammals. Anti-idiotype antibodies can be used to assess the level of a
CC   specific T cell carrying these receptors in a mammal being treated using
CC   these methods. Host cells and vectors carrying nucleic acid encoding
CC   a TCR (or individual alpha or beta chain fragment) are useful in
CC   pharmaceutical compositions to prevent or treat cancer in a mammal, e.g.
CC   lung, melanoma, ovarian, colon, brain or kidney tumours.
SQ   Sequence   10 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 10;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 EAAGIGILTV 10
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT    6
ID   W32269 standard; peptide; 10 AA.
AC   W32269;
DT   13-MAR-1998  (first entry)
DE   Tumour rejection antigen #2.
KW   Tumour rejection antigen; immunogen; TRA; cytotoxic T cell; CTL;
KW   granulyte-macrophage colony stimulating factor; GM-CSF; adjuvant.
OS   Homo sapiens.
PN   WO9728816-A1.
PD   14-AUG-1997.
PF   28-JAN-1997; U01249.
PR   09-FEB-1996; US-598909.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Jager E,  Knuth A;
DR   WPI; 97-415070/38.
PT   Composition containing immunogen and granulocyte macrophage colony
PT   stimulating factor as adjuvant - particularly for generating a
PT   cytotoxic T cell response to tumour antigens or their precursors
PS   Claim 7; Page 12; 37pp; English.
CC   This sequence represents a specifically claimed example of a tumour
CC   rejection antigen (TRA) which was used with granulocyte macrophage
CC   colony-stimulating factor (GM-CSF) as adjuvant to generate an immune,
CC   specifically cytolytic T cell (CTL), response for treatment of cancers
CC   or where cell transformation has occurred, e.g. in melanoma or dysplastic
CC   nevi. These tumour rejection antigens can also be used diagnostically (if
CC   they can induce CTL or antibodies specific for the antigens then this
CC   indicates presence of the antigen in the patient). GM-CSF provokes, or
CC   increases, immune response to the tumour rejection antigens.
SQ   Sequence   10 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 10;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 EAAGIGILTV 10
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT    7
ID   W54809 standard; peptide; 10 AA.
AC   W54809;
DT   29-SEP-1998  (first entry)
DE   Peptide 1 from Mart-1/Melan-A.
KW   Mannose; antigen; antigen-presenting cell; mannosylated peptide; T cell;
KW   vaccine; treatment.
OS   Synthetic.
PN   WO9813378-A1.
PD   02-APR-1998.
PF   25-SEP-1997; NL0536.
PR   26-SEP-1996; EP-202701.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Drijfhout JW, Koning F;
DR   WPI; 98-230631/20.
PT   Increasing uptake and presentation of antigen(s) - by adding mannose
PT   residue(s) to antigen for increasing T cell response, useful in,
PT   e.g. vaccines against viral infection(s)
PS   Disclosure; Page 25; 47pp; English.
CC   The peptides W54559-W54809 are examples of peptides to which at least 1
CC   (preferably 2) mannose can be attached to increase their uptake as
CC   antigens by antigen-presenting cells.  Uptake of agonist mannosylated
CC   peptides will increase the T cell response, whereas uptake of antagonist
CC   peptides blocks the T cell response.  Blocking binding of immunogenic
CC   autoantigens can be used in treatment of type I diabetes, rheumatoid
CC   arthritis, graft rejection etc., also to induce T-cell non-
CC   responsiveness.  Vaccines containing mannosylated antigen are used to
CC   prevent or treat infections by, e.g. bacteria, viruses, fungi, helminths
CC   and parasites.
SQ   Sequence   10 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 10;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 EAAGIGILTV 10
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT    8
ID   W22039 standard; peptide; 10 AA.
AC   W22039;
DT   20-FEB-1998  (first entry)
DE   Antigenic MART-1 peptide M10-3.
KW   Antigenic peptide; human papillomavirus; MART-1; M10-3; MAGE gene;
KW   human immunodeficiency virus; cancer antigen; tyrosinase; signal protein;
KW   anthrax lethal factor; LF; toxin; cationic fusion peptide; translocation;
KW   gene therapy; polycationic affinity handle; therapeutic protein; LFN.
OS   Homo sapiens.
PN   WO9723236-A1.
PD   03-JUL-1997.
PF   13-DEC-1996; U20463.
PR   07-JUN-1996; US-019275.
PR   13-DEC-1995; US-008518.
PA   (HARD ) HARVARD COLLEGE.
PI   Ballard JD,  Blanke SR,  Collier RJ,  Lyszak EL,  Milne JC;
PI   Starnbach MN;
DR   WPI; 97-350782/32.
PT   Introducing therapeutic proteins, especially antigens, into cells -
PT   using toxin molecules and/or polycationic handles for delivery
PS   Claim 15; Page 37; 67pp; English.
CC   This is the antigenic MART-1 peptide M10-3. This antigenic compound
CC   can be introduced into the cytoplasm of a cell by a new method where
CC   the cell is contacted with a fusion molecule comprising a delivery
CC   molecule. The delivery molecule can either be a polycationic affinity
CC   handle, LFN (the protective antigen binding domain of anthrax lethal
CC   factor) or a toxin delivery molecule related to LFN. The antigenic
```

```
CC  compound is linked to either of the delivery molecules by a covalent
CC  bond. The B moiety of a toxin enhances delivery of the antigenic compound
CC  into a cell. The anthrax toxin system of the invention eliminates the
CC  need to generate fusion proteins with a toxin B moiety, which alleviates
CC  problems associated with incorrect folding of lengthy fusion proteins.
CC  Small cationic fusion peptides substituted for LFN may reduce the
CC  possibility of steric interference with the biological activity of the
CC  translocated protein. The method is used for the introduction of
CC  antigens, e.g. MHC class I antigens or any other therapeutic protein,
CC  e.g. toxin molecules, apoptosis-inducing molecules or signalling
CC  proteins into the cells.
SQ  Sequence  10 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 10;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches   10;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 EAAGIGILTV 10
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT   9
ID  R84197 standard; Peptide; 10 AA.
AC  R84197;
DT  20-APR-1996  (first entry)
DE  MART-1 melanoma antigen immunogenic peptide M10-3 derivative.
KW  MART-1; M10-3; melanoma antigen recognised by T-cells; melanoma;
KW  metastatic melanoma; tumour-associated antigen;
KW  immunogenic peptide; diagnosis; prognosis; prophylaxis;
KW  therapy; vaccine.
OS  Synthetic.
PN  WO9529193-A2.
PD  02-NOV-1995.
PF  21-APR-1995; U05063.
PR  22-APR-1994; US-231565.
PR  05-APR-1995; US-417174.
PA  (USSH ) US SEC DEPT HEALTH.
PI  Kawakami Y,  Rosenberg SA;
DR  WPI; 95-382963/49.
PT  DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT  vectors, host cells and antibodies, used to detect, treat and
PT  immunise animal against melanoma.
PS  Claim 12; Page 122; 184pp; English.
CC  Immunogenic peptide M10-3 is a derivative of peptide M9-2 (R84196)
CC  which is based on the melanoma antigen (MART-1) (see R84212).
CC  M9-2 may be modified to improve immunogenicity (see R84783-R84800)
CC  and used in medicaments for the treatment or prevention (by
CC  immunization) of melanoma.  Antibodies against MART-1 and its
CC  immunogenic peptides may be used in the detection and isolation of
CC  MART-1 from a sample, the detection of which is indicative of a
CC  disease state (melanoma or metastatic melanoma).
CC  See also R84198
SQ  Sequence  10 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 10;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches   10;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 EAAGIGILTV 10
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT   10
ID  R84212 standard; Protein; 118 AA.
AC  R84212;
DT  20-APR-1996  (first entry)
DE  MART-1 melanoma antigen.
KW  MART-1; melanoma antigen recognised by T-cell; melanoma;
KW  metastatic melanoma; tumour-associated antigen; immunogen;
KW  diagnosis; prognosis; prophylaxis; therapy; vaccine.
OS  Mammalian.
FH  Key             Location/Qualifiers
FT  region          27. .47
FT                  /note= "hydrophobic region"
PN  WO9529193-A2.
PD  02-NOV-1995.
PF  21-APR-1995; U05063.
PR  22-APR-1994; US-231565.
PR  05-APR-1995; US-417174.
PA  (USSH ) US SEC DEPT HEALTH.
PI  Kawakami Y,  Rosenberg SA;
DR  WPI; 95-382963/49.
DR  N-PSDB; T02714.
PT  DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT  vectors, host cells and antibodies, used to detect, treat and
PT  immunise animal against melanoma.
PS  Claim 11; Page 117; 184pp; English.
CC  The melanoma antigen (MART-1) is produced by recombinant DNA
CC  methods, i.e. preferably using a baculovirus vector for expression
CC  in insect cell cultures.  MART-1 protein is a source of immunogenic
CC  peptides (see R84196 for peptide M9-2) which are optionally modified
CC  (see R84783-R84800) and used in medicaments for the treatment or
CC  prevention (by immunization) of melanoma.  Antibodies against MART-1
CC  and its immunogenic peptides may be used in the detection and
CC  isolation of MART-1 from a sample, the detection of which is
CC  indicative of a disease state (melanoma or metastatic melanoma).
SQ  Sequence  118 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches   10;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     26 EAAGIGILTV 35
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT   11
ID  R63158 standard; Protein; 118 AA.
AC  R63158;
DT  26-MAY-1995  (first entry)
DE  Tumour rejection antigen precursor.
KW  Tumour rejection antigen; precursor; HLA-A2 molecule; tyrosinase;
KW  isolation; melanoma; cell line; LB-39-MEL; diagnosis; vaccine;
KW  therapy.
OS  Homo sapiens.
PN  WO9421126-A.
PD  29-SEP-1994.
PF  09-MAR-1994; U02487.
PR  18-MAR-1993; US-032978.
PA  (LUDW-) LUDWIG INST CANCER RES.
PI  Boon-Falleur T,  Brichard V,  De Plaen E,  Traversari C;
PI  Van Pel A, Wolfel T;
DR  WPI; 94-316544/39.
DR  N-PSDB; Q76370.
PT  Nucleic acid coding for a tumour rejection antigen precursor - is
PT  used for developing prods. for diagnosis or treatment of expression
PT  related disorders, partic. melanoma
PS  Claim 5; Page 14; 26pp; English.
CC  This sequence represents the tumour rejection antigen precursor which is
CC  processed to a tumour rejection antigen presented by HLA-A2 molecules.
CC  The tumour rejection antigen is not related to tyrosinase.  The cDNA
CC  encoding this sequence was isolated from the melanoma cell line,
CC  LB-39-MEL.  The tumour rejection antigen may be used for diagnosis or
CC  in vaccines or for therapy of disorders characterised by the expression
CC  of the tumour rejection antigen precursor, particularly melanoma.
SQ  Sequence  118 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches   10;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     26 EAAGIGILTV 35
          ||||||||||
```

```
Qy      1 EAAGIGILTV 10


RESULT   12
ID   W83134 standard; Protein; 118 AA.
AC   W83134;
DT   04-FEB-1999  (first entry)
DE   Human tumour rejection antigen precursor.
KW   Human; tumour rejection antigen precursor; human leukocyte antigen;
KW   TRAP; HLA; cancer; melanoma.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 2
FT                   /note= "encoded by CGA"
FT   Misc_difference 17
FT                   /note= "encoded by GAC"
PN   US5837476-A.
PD   17-NOV-1998.
PF   16-JAN-1998; 007966.
PR   03-MAR-1995; US-398409.
PR   16-JAN-1998; US-007966.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon-Falleur T, Brichard V, De Plaen E, Traversari C,
PI   Van Pel A, Woelfelt;
DR   WPI; 99-043967/04.
DR   N-PSDB; V70150.
PT   Use of a tumour rejection antigen precursor - as a marker for
PT   diagnosing a disorder characterised by expression of a tumour
PT   rejection antigen precursor which is not tyrosinase
PS   Claim 1; Column 7-9; 11pp; English.
CC   A method has been developed for the diagnosis of a disorder which is
CC   characterised by the expression of a tumour rejection antigen precursor
CC   (TRAP) which is not tyrosinase, and which is processed to a TRA which
CC   forms a complex with an HLA-A2 molecule. The present sequence represents
CC   the TRAP for use in the present invention. The method comprises
CC   contacting a sample from a subject with an agent specific for the
CC   complex and determining the interaction between the complex and the
CC   agent as a determination of the disorder. TRAP can be used for the
CC   diagnosis and treatment of disorders characterised by the expression
CC   of the TRAP molecules such as cancers, particularly melanoma.
SQ   Sequence   118 AA;

  Query Match          100.0%;  Score 62;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 2.79e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     26 EAAGIGILTV 35
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT   13
ID   W98936 standard; peptide; 10 AA.
AC   W98936;
DT   06-MAY-1999  (first entry)
DE   Human leukocyte antigen A2 molecule binding peptide SEQ ID NO:24.
KW   Human leukocyte antigen; HLA; HLA-A2 binding peptide; T cell;
KW   cytolytic T cell; CTL.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9858951-A1.
PD   30-DEC-1998.
PF   18-JUN-1998; U12879.
PR   16-APR-1998; US-061388.
PR   23-JUN-1997; US-880963.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Cerottini J, Romero P, Valmori D;
DR   WPI; 99-105609/09.
PT   New decamer peptides which bind to HLA molecules - useful to
PT   identify HLA-A2 positive cells and provoke T cells
PS   Claim 13; Page 21; 45pp; English.
CC   The present invention describes peptides which bind to an HLA-A2
CC   molecule and have Val at the carboxy terminus, and either: (a) Ala, Tyr
CC   or Phe at the amino terminus, and Ala at position 2 (P1); or (b) Glu at
CC   the amino terminus, and Ala, Leu, or Met at positions 2 and 3, with the
CC   proviso that Ala is not at both positions (P2). The present sequence
CC   represents an HLA-A2 binding peptide. The peptides of the present
CC   invention are used to identify HLA-A2 positive cells, provoke T cells,
CC   and determine the presence of particular T cells including cytolytic
CC   T cells (CTLs). They provide a better target than the prior art
CC   CTL-stimulating peptide.
SQ   Sequence   10 AA;

  Query Match          95.2%;  Score 59;  DB 1;  Length 10;
  Best Local Similarity 90.0%;  Pred. No. 6.17e+00;
  Matches    9;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 EAAGIGILAV 10
          |||||||||:|
Qy      1 EAAGIGILTV 10


RESULT   14
ID   Y10567 standard; Peptide; 9 AA.
AC   Y10567;
DT   12-MAY-1999  (first entry)
DE   HLA Class I motif peptide SEQ ID NO:497.
KW   Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW   immunisation; tumour; infectious disease; immunotherapy; cancer;
KW   malignant melanoma; viral disease; hepatitis; AIDS.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9902183-A2.
PD   21-JAN-1999.
PF   10-JUL-1998; U14289.
PR   10-DEC-1997; US-988320.
PR   10-JUL-1997; CA-209815.
PA   (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI   Kuendig TM, Simard JJL;
DR   WPI; 99-120514/10.
PT   Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT   of antigen in the lymphatic system of a mammal so as to provide a
PT   sustained CTL response, used to treat, e.g. AIDS
PS   Disclosure; Page 47; 199pp; English.
CC   The present invention describes a method of inducing and/or sustaining
CC   an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC   method comprises: (a) delivering an antigen to the mammal at a level to
CC   induce an immunological CTL response in the mammal; and (b) maintaining
CC   the level of the antigen in the mammal's lymphatic system to maintain
CC   the immunologic CTL response. The method can be used for the delivery of
CC   e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC   an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC   gene antigen, or a viral antigen. They can be used for the treatment of
CC   disease such as cancer, e.g. malignant melanoma or infectious disease,
CC   e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC   to the lymphatic system provides for potent CTL stimulation that takes
CC   place in the milieu of the lymphoid organ, and it sustains stimulation
CC   that is necessary to keep CTL active, cytotoxic and recirculating
CC   through the body. Y10071 to Y10639 represent examples of peptide
CC   antigens given in the present invention.
SQ   Sequence   9 AA;

  Query Match          90.3%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 1.36e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      2 AAGIGILTV 10


RESULT   15
ID   Y00713 standard; peptide; 9 AA.
AC   Y00713;
DT   12-MAY-1999  (first entry)
DE   Tumour antigen booster peptide Melan-AMART-1 HLA-A2 #2.
```

KW Tumour antigen; booster peptide; immune response modulation; allergy;
KW immune response enhancer; tumour cell; tumour rejection antigen;
KW leukocyte antigen-presenting molecule; autoimmune disease;
KW allograft rejection.
OS Homo sapiens.
PN WO9858956-A2.
PD 30-DEC-1998.
PF 19-JUN-1998; U12894.
PR 23-JUN-1997; US-880979.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-Falleur T, Uyttenhove C, Warnier G;
DR WPI; 99-105612/09.
PT Immunization methods using viruses expressing antigen for priming
PT and booster immunizations - useful for modulating immune responses
PT against antigen, e.g. enhancing immune response against tumour cells
PT expressing tumour rejection antigens
PS Disclosure; Page 10; 33pp; English.
CC This sequence represents a tumour antigen booster peptide that can be
CC used in the method of the invention. The method is for for modulating an
CC immune response in a mammal against an antigen, and comprises:
CC (A) inducing an immune response by: (i) administering a virus containing
CC a nucleic acid molecule encoding the antigen or its precursor to generate
CC an immune response; and (ii) administering at least one booster dose
CC comprising a peptide including the antigen, in an adjuvant, in a combined
CC amount effective to enhance the initial immune response; or
CC (B) reducing an immune response as defined for (A) but using a
CC non-adjuvant with the peptide which includes the antigen, in an amount
CC effective to reduce the initial immune response. Method (A) is used to
CC enhance the immune response against tumour cells expressing tumour
CC rejection antigens, and against pathogens in subjects having human
CC leukocyte antigen-presenting molecules. Method (B) is used to reduce the
CC immune response in allergy, autoimmune disease, and allograft rejection.
CC Method (A) provides an immunisation method which, unlike prior art, is
CC not limited by the host immune response against viral vectors.
SQ Sequence 9 AA;

Query Match 90.3%; Score 56; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.36e+01;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      1 AAGIGILTV 9
          |||||||||
Qy      2 AAGIGILTV 10
```

Search completed: Fri May 5 22:07:42 2000
Job time : 33 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:10:16 2000; MasPar time 7.35 Seconds
94.341 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-17
Description: (1-10) from US09267439.pep
Perfect Score: 62
Sequence: 1 EAAGIGILTV 10

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 22.226; Variance 25.879; scale 0.859

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	53	85.5	766	10	O23161	RECEPTOR KINASE-LIKE P	7.44e-01
2	48	77.4	808	10	Q9ZT37	PUTATIVE GLUTAMATE REC	9.67e+00
3	47	75.8	165	2	P71810	HYPOTHETICAL 18.2 KD P	1.58e+01
4	47	75.8	250	2	O31597	YJBA PROTEIN.	1.58e+01
5	47	75.8	478	14	Q87090	GLYCOPROTEIN GIII.	1.58e+01
6	47	75.8	479	14	Q87089	GLYCOPROTEIN GIII.	1.58e+01
7	47	75.8	479	14	Q87091	GLYCOPROTEIN GIII.	1.58e+01
8	47	75.8	773	10	O22178	PUTATIVE RECEPTOR PROT	1.58e+01
9	47	75.8	2219	5	Q23388	ZK1067.2 PROTEIN.	1.58e+01
10	46	74.2	848	5	O18139	T26H2.7 PROTEIN.	2.57e+01
11	46	74.2	1347	2	O30426	XYLANASE.	2.57e+01
12	45	72.6	339	1	O30640	METHYLCOBAMIDE:COM MET	4.15e+01
13	45	72.6	339	1	Q48928	METHYLCOBAMIDE:COM MET	4.15e+01
14	45	72.6	339	1	Q48950	METHYLCOBALAMIN: COENZ	4.15e+01
15	45	72.6	370	8	O48172	CYTOCHROME B.	4.15e+01
16	45	72.6	420	11	O08833	BILE ACID COA: AMINO A	4.15e+01
17	45	72.6	420	11	Q63276	KAN-1.	4.15e+01
18	45	72.6	980	5	Q17592	SIMILARITY TO INSULIN-	4.15e+01
19	45	72.6	1805	11	Q63661	FISHER 344 PRE-SIALOMU	4.15e+01
20	44	71.0	98	2	Q9Z8X3	HYPOTHETICAL 10.3 KD P	6.63e+01
21	44	71.0	467	2	P94745	FLAGELLAR HOOK-ASSOCIA	6.63e+01
22	44	71.0	620	1	O29198	ABC TRANSPORTER, ATP-B	6.63e+01
23	43	69.4	91	11	O54712	ACETYLCHOLINE RECEPTOR	1.05e+02
24	43	69.4	123	2	O83551	ANTI-SIGMA F FACTOR AN	1.05e+02
25	43	69.4	190	10	Q9XIA3	F13F21.20 PROTEIN.	1.05e+02
26	43	69.4	243	2	O86207	PHOSPHOGLYCOLATE PHOSP	1.05e+02
27	43	69.4	361	2	Q9X336	PXO1-66.	1.05e+02
28	43	69.4	402	2	O05585	ARCA.	1.05e+02
29	43	69.4	407	2	Q9X5P8	CYTOCHROME P450 HYDROX	1.05e+02
30	43	69.4	461	2	O68124	HYPOTHETICAL 49.5 KD P	1.05e+02
31	43	69.4	509	2	O05457	HYPOTHETICAL 53.2 KD P	1.05e+02
32	43	69.4	520	5	P91840	W05H5.3 PROTEIN.	1.05e+02
33	43	69.4	579	2	Q9X7A9	PUTATIVE ACYLTRANSFERA	1.05e+02
34	43	69.4	580	2	O53208	PUTATIVE TRANSFERASE.	1.05e+02
35	43	69.4	589	10	Q43390	RCH2 PROTEIN.	1.05e+02
36	43	69.4	593	2	P94513	AUTOLYSIN SENSOR KINAS	1.05e+02
37	43	69.4	667	2	P71749	HYPOTHETICAL 68.3 KD P	1.05e+02
38	43	69.4	729	10	O65497	PUTATIVE SUGAR TRANSPO	1.05e+02
39	43	69.4	950	2	Q50470	PKS002C (FRAGMENT).	1.05e+02
40	43	69.4	1436	3	Q07527	HYPOTHETICAL 165.0 KD	1.05e+02
41	43	69.4	1616	2	P96285	HYPOTHETICAL 166.6 KD	1.05e+02
42	43	69.4	7576	2	Q9ZGA4	FK506 POLYKETIDE SYNTH	1.05e+02
43	42	67.7	269	2	Q9Z7I4	KDO SYNTHETASE.	1.65e+02
44	42	67.7	347	5	Q9XX84	Y102A5C.29 PROTEIN.	1.65e+02
45	42	67.7	768	5	O46043	PARG PROTEIN.	1.65e+02

ALIGNMENTS

RESULT 1
ID O23161 PRELIMINARY; PRT; 766 AA.
AC O23161;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RECEPTOR KINASE-LIKE PROTEIN (EC 2.7.1.).
GN C7A10.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., TERRYN N., VOS P., HEIJNEN L., MEWES H.W., SCHUELLER C.,
RA CHALWATZIS N.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z99707; CAB16774.1; -.
DR MENDEL; 25486; Arath;3435;25486.
DR PFAM; PF00560; LRR; 4.
DR PFAM; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
SQ SEQUENCE 766 AA; 83775 MW; C4BD7115 CRC32;

Query Match 85.5%; Score 53; DB 10; Length 766;
Best Local Similarity 70.0%; Pred. No. 7.44e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Db    335 DIAGIGILAV 344
          : |||||||:|
Qy      1 EAAGIGILTV 10
```

RESULT 2
ID Q9ZT37 PRELIMINARY; PRT; 808 AA.
AC Q9ZT37;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE GLUTAMATE RECEPTOR.
GN GLR1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99039497.
RA   LAM H.M., CHIU J., HSIEH M.H., MEISEL L., OLIVEIRA I.C., SHIN M.,
RA   CORUZZI G.;
RT   "Glutamate-receptor genes in plants.";
RL   Nature 396:125-126(1998).
DR   EMBL; AF079998; AAD09173.1; -.
KW   Receptor.
SQ   SEQUENCE   808 AA;  90518 MW;  C3554B89 CRC32;

  Query Match           77.4%;  Score 48;  DB 10;  Length 808;
  Best Local Similarity 100.0%;  Pred. No. 9.67e+00;
  Matches    7;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    519 GIGILTV 525
          |||||||
Qy      4 GIGILTV 10


RESULT    3
ID   P71810      PRELIMINARY;      PRT;   165 AA.
AC   P71810;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 18.2 KD PROTEIN.
GN   MTCY02B12.16.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   MCLEAN J., HARRIS D.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; Z81011; CAB02643.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   165 AA;  18189 MW;  BFB84C79 CRC32;

  Query Match           75.8%;  Score 47;  DB 2;  Length 165;
  Best Local Similarity 75.0%;  Pred. No. 1.58e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    128 AGIGILAI 135
          ||||||::
Qy      3 AGIGILTV 10


RESULT    4
ID   O31597      PRELIMINARY;      PRT;   250 AA.
AC   O31597;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YJBA PROTEIN.
GN   YJBA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99110; CAB12998.1; -.
SQ   SEQUENCE   250 AA;  30119 MW;  C96222FD CRC32;

  Query Match           75.8%;  Score 47;  DB 2;  Length 250;
  Best Local Similarity 66.7%;  Pred. No. 1.58e+01;
  Matches    6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     89 TDGIGILAV 97
          : |||||:|
Qy      2 AAGIGILTV 10


RESULT    5
ID   Q87090      PRELIMINARY;      PRT;   478 AA.
AC   Q87090;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIANA S;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
```

```
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49436; BAA08414.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   478 AA;  51150 MW;  D6A143B4 CRC32;

  Query Match           75.8%;  Score 47;  DB 14;  Length 478;
  Best Local Similarity  75.0%;  Pred. No. 1.58e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    455 AGIGILAI 462
          ||||||::
Qy      3 AGIGILTV 10


RESULT    6
ID   Q87089      PRELIMINARY;      PRT;   479 AA.
AC   Q87089;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YAMAGATA S-81;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49435; BAA08413.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   479 AA;  51109 MW;  A009EB9B CRC32;

  Query Match           75.8%;  Score 47;  DB 14;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 1.58e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      3 AGIGILTV 10


RESULT    7
ID   Q87091      PRELIMINARY;      PRT;   479 AA.
AC   Q87091;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIA3;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49437; BAA08415.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   479 AA;  51148 MW;  CC3EFF9A CRC32;

  Query Match           75.8%;  Score 47;  DB 14;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 1.58e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      3 AGIGILTV 10


RESULT    8
ID   O22178      PRELIMINARY;      PRT;   773 AA.
AC   O22178;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE RECEPTOR PROTEIN KINASE.
GN   T20D16.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002391; AAB87101.1; -.
DR   MENDEL; 25106; Arath;3435;25106.
DR   PFAM; PF00560; LRR; 4.
DR   PFAM; PF00069; pkinase; 1.
SQ   SEQUENCE   773 AA;  84148 MW;  83C3953B CRC32;

  Query Match           75.8%;  Score 47;  DB 10;  Length 773;
  Best Local Similarity  60.0%;  Pred. No. 1.58e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    341 DIAGIGILAL 350
          : ||||||::
Qy      1 EAAGIGILTV 10


RESULT    9
ID   Q23388      PRELIMINARY;      PRT;  2219 AA.
AC   Q23388;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   ZK1067.2 PROTEIN.
GN   ZK1067.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   THOMAS K.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z70038; CAA93884.1; -.
SQ   SEQUENCE   2219 AA;  253649 MW;  59DE8B43 CRC32;

  Query Match           75.8%;  Score 47;  DB 5;  Length 2219;
  Best Local Similarity  60.0%;  Pred. No. 1.58e+01;
```

```
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db   1443 ENTGIGFLAV 1452
          | :|||:|:|
Qy      1 EAAGIGILTV 10


RESULT   10
ID   O18139      PRELIMINARY;      PRT;   848 AA.
AC   O18139;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T26H2.7 PROTEIN.
GN   T26H2.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z82055; CAB04848.1; -.
SQ   SEQUENCE   848 AA;  98312 MW;  371853A7 CRC32;

  Query Match           74.2%;  Score 46;  DB 5;  Length 848;
  Best Local Similarity  66.7%;  Pred. No. 2.57e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    340 SASIGILTI 348
          :|:||||||:
Qy      2 AAGIGILTV 10


RESULT   11
ID   O30426      PRELIMINARY;      PRT;  1347 AA.
AC   O30426;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-AUG-1999 (TrEMBLrel. 11, Last annotation update)
DE   XYLANASE.
GN   XYNF.
OS   Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Thermoanaerobacter group; Caldicellulosiruptor.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90253140.
RA   LUTHI E., LOVE D.R., MCANULTY J., WALLACE C., CAUGHEY P.A., SAUL D.,
RA   BERGQUIST P.L.;
RT   "Cloning, sequence analysis, and expression of genes encoding xylan-
RT   degrading enzymes from the thermophile 'Caldocellum
RT   saccharolyticum'.";
RL   Appl. Environ. Microbiol. 56:1017-1024(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   TE'O V.S. JR., GIBBS M.D., SAUL D.J., BERGQUIST P.L.;
RL   Appl. Environ. Microbiol. 0:0-0(1997).
DR   EMBL; AF005383; AAB87371.1; -.
KW   Xylan degradation; Hydrolase; Glycosidase.
SQ   SEQUENCE   1347 AA;  152048 MW;  98FFDA72 CRC32;

  Query Match           74.2%;  Score 46;  DB 2;  Length 1347;
  Best Local Similarity  66.7%;  Pred. No. 2.57e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    992 NASGIGVLT 1000
          :|:|||:||
Qy      1 EAAGIGILT 9


RESULT   12
ID   O30640      PRELIMINARY;      PRT;   339 AA.
AC   O30640;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METHYLCOBAMIDE:COM METHYLTRANSFERASE ISOZYME A.
GN   MTBA.
OS   Methanosarcina barkeri.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS (DSM 800);
RX   MEDLINE; 97341199.
RA   BURKE S.A., KRZYCKI J.A.;
RT   "Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a
RT   corrinoid protein and two methyltransferases purified from
RT   Methanosarcina barkeri.";
RL   J. Biol. Chem. 272:16570-16577(1997).
DR   EMBL; AF013713; AAC38632.1; -.
DR   PFAM; PF01208; URO-D; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   339 AA;  36664 MW;  040E3CF3 CRC32;

  Query Match           72.6%;  Score 45;  DB 1;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 4.15e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      3 AGIGILTV 10


RESULT   13
ID   Q48928      PRELIMINARY;      PRT;   339 AA.
AC   Q48928;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METHYLCOBAMIDE:COM METHYLTRANSFERASE ISOZYME A.
GN   CMTA.
OS   Methanosarcina barkeri.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH;
RX   MEDLINE; 96324952.
RA   LECLERC G.M., GRAHAME D.A.;
RT   "Methylcobamide:coenzyme M methyltransferase isozymes from
RT   Methanosarcina barkeri. Physicochemical characterization, cloning,
RT   sequence analysis, and heterologous gene expression.";
RL   J. Biol. Chem. 271:18725-18731(1996).
DR   EMBL; U38919; AAC44214.1; -.
DR   PFAM; PF01208; URO-D; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   339 AA;  36708 MW;  731F945B CRC32;
```

```
  Query Match           72.6%;  Score 45;  DB 1;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 4.15e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      3 AGIGILTV 10


RESULT   14
ID   Q48950      PRELIMINARY;      PRT;   339 AA.
AC   Q48950;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METHYLCOBALAMIN: COENZYME M METHYLTRANSFERASE (ISOENZYME II).
GN   MTBA.
OS   Methanosarcina barkeri.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FUSARO (DSM 804);
RX   MEDLINE; 96184544.
RA   HARMS U., THAUER R.K.;
RT   "Methylcobalamin: coenzyme M methyltransferase isoenzymes MtaA and
RT   MtbA from Methanosarcina barkeri. Cloning, sequencing and differential
RT   transcription of the encoding genes, and functional overexpression of
RT   the mtaA gene in Escherichia coli.";
RL   Eur. J. Biochem. 235:653-659(1996).
DR   EMBL; X91894; CAA62996.1; -.
DR   PFAM; PF01208; URO-D; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   339 AA;  36761 MW;  5F6F0A9C CRC32;

  Query Match           72.6%;  Score 45;  DB 1;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 4.15e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      3 AGIGILTV 10


RESULT   15
ID   O48172      PRELIMINARY;      PRT;   370 AA.
AC   O48172;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B.
GN   COB.
OS   Polytomella sp. 'Pringsheim 198.80'.
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Polytomella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=198.80, FROM E.G. PRINGSHEIM;
RA   ANTARAMIAN A., FUNES-ARGUELLO S., VAZQUEZ-ACEVEDO M., CORIA R.,
RA   GONZALEZ-HALPHEN D.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U87396; AAC24896.1; -.
DR   MENDEL; 23585; Pols;cob;23585.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   Mitochondrion.
SQ   SEQUENCE   370 AA;  41226 MW;  5D617081 CRC32;

  Query Match           72.6%;  Score 45;  DB 8;  Length 370;
  Best Local Similarity  85.7%;  Pred. No. 4.15e+01;
  Matches     6;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    288 GIGILAV 294
          |||||:|
Qy      4 GIGILTV 10


Search completed: Fri May  5 22:11:45 2000
Job time : 89 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:09:18 2000; MasPar time 3.01 Seconds
99.300 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-17
Description: (1-10) from US09267439.pep
Perfect Score: 62
Sequence: 1 EAAGIGILTV 10

Scoring table: PAM 150
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 23.268; Variance 24.186; scale 0.962

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	62	100.0	118	1	MAR1_HUMAN	MELANOMA ANTIGEN RECOG	8.83e-04
2	48	77.4	332	1	ACOA_ALCEU	ACETOIN:2,6-DICHLOROPH	3.08e+00
3	47	75.8	101	1	ATPL_SULAC	MEMBRANE-ASSOCIATED AT	5.22e+00
4	47	75.8	110	1	VCAD_LAMBD	HEAD DECORATION PROTEI	5.22e+00
5	47	75.8	384	1	PQQE_METEX	COENZYME PQQ SYNTHESIS	5.22e+00
6	47	75.8	479	1	VGLC_PRVIF	GLYCOPROTEIN GIII PREC	5.22e+00
7	47	75.8	752	1	METE_ECOLI	5-METHYLTETRAHYDROPTER	5.22e+00
8	46	74.2	231	1	YSA1_YEAST	YSA1 PROTEIN.	8.77e+00
9	46	74.2	271	1	YK23_YEAST	HYPOTHETICAL 31.0 KD P	8.77e+00
10	45	72.6	201	1	Y760_PYRHO	HYPOTHETICAL PROTEIN P	1.46e+01
11	45	72.6	291	1	Y4TQ_RHISN	PROBABLE PEPTIDE ABC T	1.46e+01
12	44	71.0	394	1	FTSZ_AZOVI	CELL DIVISION PROTEIN	2.41e+01
13	44	71.0	456	1	GLMU_ECOLI	UDP-N-ACETYLGLUCOSAMIN	2.41e+01
14	44	71.0	467	1	FLID_ECOLI	FLAGELLAR HOOK-ASSOCIA	2.41e+01
15	44	71.0	635	1	XYND_PAEPO	ENDO-1,4-BETA-XYLANASE	2.41e+01
16	43	69.4	132	1	ATPE_ARATH	ATP SYNTHASE EPSILON C	3.94e+01
17	43	69.4	207	1	RL4_BACSU	50S RIBOSOMAL PROTEIN	3.94e+01
18	43	69.4	345	1	IAP_ECOLI	ALKALINE PHOSPHATASE I	3.94e+01
19	43	69.4	404	1	SGAA_HYPME	SERINE--GLYOXYLATE AMI	3.94e+01
20	43	69.4	461	1	YXCC_BACSU	HYPOTHETICAL METABOLIT	3.94e+01
21	43	69.4	493	1	ACHE_MOUSE	ACETYLCHOLINE RECEPTOR	3.94e+01
22	43	69.4	590	1	CHL1_ARATH	NITRATE/CHLORATE TRANS	3.94e+01
23	43	69.4	611	1	YD3M_HERAU	HYPOTHETICAL 68.4 KD P	3.94e+01
24	43	69.4	651	1	BGLR_CANFA	BETA-GLUCURONIDASE PRE	3.94e+01
25	43	69.4	675	1	NUAM_ACACA	NADH DEHYDROGENASE, SU	3.94e+01
26	43	69.4	1217	1	EGF_MOUSE	PRO-EPIDERMAL GROWTH F	3.94e+01
27	43	69.4	1325	1	YDEK_ECOLI	HYPOTHETICAL 136.5 KD	3.94e+01
28	42	67.7	111	1	YPFE_ECOLI	HYPOTHETICAL 11.6 KD P	6.37e+01
29	42	67.7	216	1	FLA2_METVO	FLAGELLIN B2 PRECURSOR	6.37e+01
30	42	67.7	222	1	FLA2_METVA	FLAGELLIN B2 PRECURSOR	6.37e+01
31	42	67.7	251	1	YJFQ_ECOLI	HYPOTHETICAL TRANSCRIP	6.37e+01
32	42	67.7	308	1	MENA_HAEIN	1,4-DIHYDROXY-2-NAPHTH	6.37e+01
33	42	67.7	321	1	YHBE_ECOLI	HYPOTHETICAL 35.0 KD P	6.37e+01
34	42	67.7	325	1	RCEM_CHRVI	REACTION CENTER PROTEI	6.37e+01
35	42	67.7	345	1	HRCA_STRMU	HEAT-INDUCIBLE TRANSCR	6.37e+01
36	42	67.7	440	1	UGTC_CAEEL	PUTATIVE UDP-GLUCURONO	6.37e+01
37	42	67.7	461	1	THDF_HAEIN	POSSIBLE THIOPHENE AND	6.37e+01
38	42	67.7	501	1	LYSI_CORGL	L-LYSINE TRANSPORT PRO	6.37e+01
39	42	67.7	503	1	SECD_HELPY	PROTEIN-EXPORT MEMBRAN	6.37e+01
40	42	67.7	526	1	SECD_HELPJ	PROTEIN-EXPORT MEMBRAN	6.37e+01
41	42	67.7	530	1	AIP2_YEAST	ACTIN INTERACTING PROT	6.37e+01
42	42	67.7	659	1	YYBT_BACSU	HYPOTHETICAL 74.3 KD P	6.37e+01
43	42	67.7	661	1	BAIH_EUBSP	NADH-DEPENDENT FLAVIN	6.37e+01
44	42	67.7	885	1	YDGH_BACSU	PUTATIVE MEMBRANE PROT	6.37e+01
45	42	67.7	1530	1	BFR1_SCHPO	BREFELDIN A RESISTANCE	6.37e+01

ALIGNMENTS

```
RESULT    1
ID   MAR1_HUMAN     STANDARD;      PRT;   118 AA.
AC   Q16655;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MELANOMA ANTIGEN RECOGNIZED BY T-CELLS 1 (MART-1) (MELAN-A PROTEIN)
DE   (ANTIGEN SK29-AA) (ANTIGEN LB39-AA).
GN   MLANA OR MART1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RX   MEDLINE; 94224770.
RA   KAWAKAMI Y., ELIYAHU S., DELGADO C.H., ROBBINS P.F., RIVOLTINI L.,
RA   TOPALIAN S.L., MIKI T., ROSENBERG S.A.;
RT   "Cloning of the gene coding for a shared human melanoma antigen
RT   recognized by autologous T cells infiltrating into tumor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3515-3519(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94275389.
RA   COULIE P.G., BRICHARD V., VAN PEL A., WOELFEL T., SCHNEIDER J.,
RA   TRAVERSARI C., MATTEI S., DE PLAEN E., LURQUIN C., SZIKORA J.-P.,
RA   RENAULD J.-C., BOON T.;
RT   "A new gene coding for a differentiation antigen recognized by
RT   autologous cytolytic T lymphocytes on HLA-A2 melanomas.";
RL   J. Exp. Med. 180:35-42(1994).
CC   -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO MELANOMA AND
CC       MELANOCYTE CELL LINES AND RETINA.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U06452; AAA19238.1; -.
DR   EMBL; U06654; AAA20389.1; -.
KW   Antigen; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
SQ   SEQUENCE   118 AA;  13157 MW;  DFE2CFC6 CRC32;
```

```
  Query Match          100.0%;  Score 62;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 8.83e-04;
  Matches   10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     26 EAAGIGILTV 35
          ||||||||||
Qy      1 EAAGIGILTV 10


RESULT    2
ID   ACOA_ALCEU     STANDARD;      PRT;   332 AA.
AC   P27745;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   ACETOIN:2,6-DICHLOROPHENOLINDOPHENOL OXIDOREDUCTASE ALPHA
DE   SUBUNIT (EC 1.1.1.-) (ACETOIN:DCPIP OXIDOREDUCTASE-ALPHA)
DE   (AO:DCPIP OR).
GN   ACOA.
OS   Alcaligenes eutrophus.
OC   Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC   Ralstonia.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.
RC   STRAIN=H16;
RX   MEDLINE; 91286190.
RA   PRIEFERT H., HEIN S., KRUEGER N., ZEH K., SCHMIDT B., STEINBUECHEL A.;
RT   "Identification and molecular characterization of the Alcaligenes
RT   eutrophus H16 aco operon genes involved in acetoin catabolism.";
RL   J. Bacteriol. 173:4056-4071(1991).
CC   -!- FUNCTION: CATALYZES THE 2,6-DICHLOROPHENOLINDOPHENOL-DEPENDENT
CC       CLEAVAGE OF ACETOIN INTO ACETATE AND ACETALDEHYDE, IN VITRO. THE
CC       ALPHA SUBUNIT IS PROBABLY THE CATALYTIC SUBUNIT OF THE ENZYME.
CC   -!- COFACTOR: THIAMINE PYROPHOSPHATE.
CC   -!- PATHWAY: ACETOIN CATABOLISM.
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
CC   -!- SIMILARITY: TO THE ALPHA SUBUNITS OF 2-OXO-ACID DEHYDROGENASE
CC       COMPONENTS OF VARIOUS MULTIENZYME COMPLEXES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M66060; AAA21948.1; -.
DR   PIR; B42462; DEALXE.
DR   PFAM; PF00676; E1_dehydrog; 1.
KW   Oxidoreductase; Flavoprotein; Thiamine pyrophosphate.
FT   INIT_MET      0      0
FT   BINDING     173    173       THIAMINE PYROPHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   332 AA;  35243 MW;  3322DA8E CRC32;

  Query Match          77.4%;  Score 48;  DB 1;  Length 332;
  Best Local Similarity 87.5%;  Pred. No. 3.08e+00;
  Matches    7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     55 EAAGVGIL 62
          ||||:|||
Qy      1 EAAGIGIL 8


RESULT    3
ID   ATPL_SULAC     STANDARD;      PRT;   101 AA.
AC   P23040;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MEMBRANE-ASSOCIATED ATPASE C CHAIN (EC 3.6.1.34) (SUL-ATPASE
DE   PROTEOLIPID CHAIN).
GN   ATPP.
OS   Sulfolobus acidocaldarius.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 89214142.
RA   DENDA K., KONISHI J., OSHIMA T., DATE T., YOSHIDA M.;
RT   "A gene encoding the proteolipid subunit of Sulfolobus acidocaldarius
RT   ATPase complex.";
RL   J. Biol. Chem. 264:7119-7121(1989).
CC   -!- FUNCTION: THE C CHAIN IS A PROTEOLIPID, AND ONE OF THE MEMBRANOUS
CC       SUBUNITS OF THE THE NONENZYMATIC COMPONENT OF THE SUL-ATPASE
CC       COMPLEX.
CC   -!- SUBUNIT: SUL-ATPASE IS COMPOSED OF SIX (OR FIVE ?) SUBUNITS:
CC       ALPHA, BETA, DELTA, GAMMA, C (PROTEOLIPID), AND POSSIBLY EPSILON.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE V-ATPASE PROTEOLIPID SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04740; AAA72703.1; -.
DR   PIR; A33351; A33351.
DR   HSSP; P00138; 1CGN.
DR   PFAM; PF00137; ATP-synt_C; 1.
KW   Hydrogen ion transport; Lipid-binding; Transmembrane.
FT   TRANSMEM      5     25       POTENTIAL.
FT   TRANSMEM     37     57       POTENTIAL.
FT   TRANSMEM     75     95       POTENTIAL.
SQ   SEQUENCE   101 AA;  10362 MW;  1DC8C74D CRC32;

  Query Match          75.8%;  Score 47;  DB 1;  Length 101;
  Best Local Similarity 87.5%;  Pred. No. 5.22e+00;
  Matches    7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 AAGIGVLT 66
          |||||:||
Qy      2 AAGIGILT 9


RESULT    4
ID   VCAD_LAMBD     STANDARD;      PRT;   110 AA.
AC   P03712;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   HEAD DECORATION PROTEIN (GPD) (MAJOR CAPSID PROTEIN D).
GN   D.
OS   Bacteriophage lambda.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC   Lambda phage group.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83189071.
RA   SANGER F., COULSON A.R., HONG G.F., HILL D.F., PETERSEN G.B.;
RT   "Nucleotide sequence of bacteriophage lambda DNA.";
RL   J. Mol. Biol. 162:729-773(1982).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 84207913.
RA   WITKIEWICZ H., SCHWEIGER M.;
RT   "The head protein D of bacterial virus lambda is related to
RT   eukaryotic chromosomal proteins.";
RL   EMBO J. 1:1559-1564(1982).
CC   -!- FUNCTION: STABILIZES THE HEAD SHELL FOLLOWING THE REARRANGEMENT
CC       OF THE GPE SUBUNITS OF THE HEAD SHELL LATTICE THAT ACCOMPANIES
CC       EXPANSION OF THE HEAD. THERE ARE APPROXIMATELY 420 COPIES OF
CC       PROTEIN D PER MATURE PHAGE.
CC   -!- SIMILARITY: TO BACTERIOPHAGE 21 HEAD DECORATION PROTEIN.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J02459; AAA96539.1; -.
DR   PIR; A04334; VHBPDL.
DR   PIR; A23206; A23206.
KW   Coat protein.
SQ   SEQUENCE   110 AA;  11572 MW;  FDD50011 CRC32;

  Query Match           75.8%;  Score 47;  DB 1;  Length 110;
  Best Local Similarity  50.0%;  Pred. No. 5.22e+00;
  Matches     5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db     55 DGAAVGILAV 64
          ::|::|||:|
Qy      1 EAAGIGILTV 10


RESULT    5
ID   PQQE_METEX     STANDARD;      PRT;   384 AA.
AC   P71517;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   COENZYME PQQ SYNTHESIS PROTEIN E.
GN   PQQE.
OS   Methylobacterium extorquens.
OC   Bacteria; Proteobacteria; alpha subdivision; Methylobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AM1 / NCIB 9133;
RX   MEDLINE; 97195805.
RA   TOYAMA H., CHISTOSERDOVA L., LIDSTROM M.E.;
RT   "Sequence analysis of pqq genes required for biosynthesis of
RT   pyrroloquinoline quinone in Methylobacterium extorquens AM1 and the
RT   purification of a biosynthetic intermediate.";
RL   Microbiology 143:595-602(1997).
CC   -!- FUNCTION: REQUIRED FOR COENZYME PYRROLO-QUINOLINE-QUINONE (PQQ)
CC       BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE MOAA/NIFB/PQQE FAMILY OF PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U72662; AAB58898.1; -.
DR   PROSITE; PS01305; MOAA_NIFB_PQQE; 1.
DR   PFAM; PF01444; MoaA_NifB_PqqE; 1.
KW   PQQ; Iron-sulfur.
FT   METAL        28     28       IRON-SULFUR (POTENTIAL).
FT   METAL        32     32       IRON-SULFUR (POTENTIAL).
FT   METAL        35     35       IRON-SULFUR (POTENTIAL).
SQ   SEQUENCE   384 AA;  41714 MW;  7BD3BE8C CRC32;

  Query Match           75.8%;  Score 47;  DB 1;  Length 384;
  Best Local Similarity  70.0%;  Pred. No. 5.22e+00;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     59 EAAGLGVLHV 68
          ||||:|:| |
Qy      1 EAAGIGILTV 10


RESULT    6
ID   VGLC_PRVIF     STANDARD;      PRT;   479 AA.
AC   P06024;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   GLYCOPROTEIN GIII PRECURSOR.
OS   Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86200375.
RA   ROBBINS A.K., WATSON R.J., WHEALY M.E., HAYS W.W., ENQUIST L.W.;
RT   "Characterization of a pseudorabies virus glycoprotein gene with
RT   homology to herpes simplex virus type 1 and type 2 glycoprotein C.";
RL   J. Virol. 58:339-347(1986).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN C.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN GENE SUPERFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M12778; AAA47464.1; -.
DR   PIR; A26097; VGBEPB.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    479       GLYCOPROTEIN GIII.
FT   CARBOHYD     40     40       POTENTIAL.
FT   CARBOHYD     84     84       POTENTIAL.
FT   CARBOHYD    169    169       POTENTIAL.
FT   CARBOHYD    192    192       POTENTIAL.
FT   CARBOHYD    220    220       POTENTIAL.
FT   CARBOHYD    228    228       POTENTIAL.
FT   CARBOHYD    285    285       POTENTIAL.
FT   CARBOHYD    302    302       POTENTIAL.
SQ   SEQUENCE   479 AA;  51206 MW;  42EE5703 CRC32;

  Query Match           75.8%;  Score 47;  DB 1;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 5.22e+00;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      3 AGIGILTV 10


RESULT    7
ID   METE_ECOLI     STANDARD;      PRT;   752 AA.
AC   P25665;
DT   01-MAY-1992 (Rel. 22, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE
DE   (EC 2.1.1.14) (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME)
DE   (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE).
GN   METE.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 92358234.
RA   DANIELS D.L., PLUNKETT G. III, BURLAND V.D., BLATTNER F.R.;
RT   "Analysis of the Escherichia coli genome: DNA sequence of the region
RT   from 84.5 to 86.5 minutes.";
RL   Science 257:771-778(1992).
```

```
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 93347969.
RA   PLUNKETT G. III, BURLAND V., DANIELS D.L., BLATTNER F.R.;
RT   "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT   region from 87.2 to 89.2 minutes.";
RL   Nucleic Acids Res. 21:3391-3398(1993).
RN   [3]
RP   REVISIONS TO 604 AND 658.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=K12 / DH5-ALPHA;
RX   MEDLINE; 92329421.
RA   GONZALEZ J.C., BANERJEE R.V., HUANG S., SUMNER J.S., MATTHEWS R.G.;
RT   "Comparison of cobalamin-independent and cobalamin-dependent
RT   methionine synthases from Escherichia coli: two solutions to the same
RT   chemical problem.";
RL   Biochemistry 31:6045-6056(1992).
RN   [5]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE; 89098936.
RA   MAXON M.E., REDFIELD B., CAI X.-Y., SHOEMAN R., FUJITA K., FISHER W.,
RA   STAUFFER G., WEISSBACH H., BROT N.;
RT   "Regulation of methionine synthesis in Escherichia coli: effect of
RT   the MetR protein on the expression of the metE and metR genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:85-89(1989).
RN   [6]
RP   SEQUENCE OF 1-12.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE; 97443975.
RA   LINK A.J., ROBISON K., CHURCH G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [7]
RP   CHARACTERIZATION, AND MUTAGENESIS OF CYS-725.
RX   MEDLINE; 96420456.
RA   GONZALEZ J.C., PEARISO K., PENNER-HAHN J.E., MATTHEWS R.G.;
RT   "Cobalamin-independent methionine synthase from Escherichia coli: a
RT   zinc metalloenzyme.";
RL   Biochemistry 35:12228-12234(1996).
CC   -!- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-
CC       METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE
CC       FORMATION.
CC   -!- CATALYTIC ACTIVITY: 5-METHYLTETRAHYDROPTEROYLTRI-L-GLUTAMATE +
CC       L-HOMOCYSTEINE = TETRAHYDROPTEROYLTRI-L-GLUTAMATE + L-METHIONINE.
CC   -!- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC   -!- COFACTOR: ZINC; BINDS ONE MOLE PER SUBUNIT.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: HAS AN ABSOLUTE REQUIREMENT FOR A POLYGLUTAMYLATED
CC       FOLATE AS SUBSTRATE. ITS ACTIVITY DEPENDS ON PHOSPHATE ANIONS AND
CC       DIVALENT CATIONS.
CC   -!- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC       SYNTHASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M87049; AAA67625.1; -.
DR   EMBL; M87625; AAA23544.1; -.
DR   EMBL; AE000458; AAC76832.1; -.
DR   EMBL; J04155; AAA24160.1; -.
DR   PIR; S30719; S30719.
DR   PIR; A42863; A42863.
DR   ECO2DBASE; F088.0; 6TH EDITION.
DR   ECOGENE; EG10584; METE.
KW   Transferase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat.
FT   INIT_MET      0      0
FT   REPEAT        1    369       APPROXIMATE.
FT   REPEAT      370    752       APPROXIMATE.
FT   METAL       725    725       ZINC.
FT   MUTAGEN     725    725       C->S: LOSS OF ACTIVITY.
FT   CONFLICT    362    362       L -> V (IN REF. 4).
FT   CONFLICT    604    604       E -> Q (IN REF. 2).
FT   CONFLICT    658    658       A -> R (IN REF. 2).
SQ   SEQUENCE   752 AA;  84542 MW;  BC7C6078 CRC32;

  Query Match           75.8%;  Score 47;  DB 1;  Length 752;
  Best Local Similarity  70.0%;  Pred. No. 5.22e+00;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    588 EAAGIGIIQI 597
          |||||||: :
Qy      1 EAAGIGILTV 10


RESULT    8
ID   YSA1_YEAST     STANDARD;      PRT;   231 AA.
AC   Q01976;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   YSA1 PROTEIN.
GN   YSA1 OR YBR111C OR YBR0907.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 95208357.
RA   MANNHAUPT G., STUCKA R., EHNLE S., VETTER I., FELDMANN H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
RN   [2]
RP   SEQUENCE OF 1-47 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 92327848.
RA   MANNHAUPT G., STUCKA R., EHNLE S., VETTER I., FELDMANN H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
CC   -!- SIMILARITY: STRONG, TO B.SUBTILIS YQKG.
CC   -!- SIMILARITY: TO PROTEINS WITH A CORE MUTT DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35980; CAA85068.1; -.
DR   EMBL; X78993; CAA55614.1; -.
DR   EMBL; X66247; CAA46972.1; -.
DR   PIR; S44691; S44691.
DR   SGD; L0002551; YSA1.
DR   PROSITE; PS00893; MUTT; 1.
DR   PFAM; PF00293; mutT; 1.
FT   DOMAIN      112    145       MUTT-LIKE.
```

```
SQ   SEQUENCE   231 AA;  26087 MW;  49A2D6CB CRC32;

  Query Match           74.2%;  Score 46;  DB 1;  Length 231;
  Best Local Similarity  85.7%;  Pred. No. 8.77e+00;
  Matches     6;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     79 GIGILTI 85
          ||||||:
Qy      4 GIGILTV 10


RESULT    9
ID   YK23_YEAST     STANDARD;      PRT;   271 AA.
AC   P36136;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   HYPOTHETICAL 31.0 KD PROTEIN IN GAP1-NAP1 INTERGENIC REGION.
GN   YKR043C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   URRESTARAZU L.A., JAUNIAUX J.-C.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z28268; CAA82119.1; -.
DR   PIR; S38115; S38115.
DR   PFAM; PF00300; PGAM; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   271 AA;  31022 MW;  F8A036A8 CRC32;

  Query Match           74.2%;  Score 46;  DB 1;  Length 271;
  Best Local Similarity  55.6%;  Pred. No. 8.77e+00;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    233 DAGGIGVLS 241
          :|:|||:|:
Qy      1 EAAGIGILT 9


RESULT   10
ID   Y760_PYRHO     STANDARD;      PRT;   201 AA.
AC   O58499;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN PH0760.
GN   PH0760 OR PHCI026.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE UPF0056 (MARC) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000003; BAA29851.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM    111    131       POTENTIAL.
FT   TRANSMEM    140    160       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
SQ   SEQUENCE   201 AA;  21592 MW;  97675186 CRC32;

  Query Match           72.6%;  Score 45;  DB 1;  Length 201;
  Best Local Similarity  60.0%;  Pred. No. 1.46e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    137 ESPGIVILTI 146
          |::|| |||:
Qy      1 EAAGIGILTV 10


RESULT   11
ID   Y4TQ_RHISN     STANDARD;      PRT;   291 AA.
AC   Q53192;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROBABLE PEPTIDE ABC TRANSPORTER PERMEASE PROTEIN Y4TQ.
GN   Y4TQ.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97305956.
RA   FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA   PERRET X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96389014.
RA   FREIBERG C., PERRET X., BROUGHTON W.J., ROSENTHAL A.;
RT   "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.
RT   NGR234 using dye terminators and a thermostable 'sequenase': a
RT   beginning.";
RL   Genome Res. 6:590-600(1996).
CC   -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM Y4TOPQRS FOR A PEPTIDE. PROBABLY RESPONSIBLE FOR THE
CC       TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPPBC
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; Z68203; CAA92399.1; -.
DR   EMBL; AE000098; AAB91870.1; -.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
DR   PFAM; PF00528; BPD_transp; 1.
KW   Hypothetical protein; Transport; Amino-acid transport; Transmembrane;
KW   Inner membrane; Plasmid.
FT   TRANSMEM     28     48       POTENTIAL.
FT   TRANSMEM     92    112       POTENTIAL.
FT   TRANSMEM    137    157       POTENTIAL.
FT   TRANSMEM    213    233       POTENTIAL.
FT   TRANSMEM    249    269       POTENTIAL.
SQ   SEQUENCE   291 AA;  30910 MW;  3263271E CRC32;

  Query Match           72.6%;  Score 45;  DB 1;  Length 291;
  Best Local Similarity  66.7%;  Pred. No. 1.46e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    147 GPGIGILIV 155
          ::|||||  |
Qy      2 AAGIGILTV 10


RESULT   12
ID   FTSZ_AZOVI     STANDARD;      PRT;   394 AA.
AC   P77817;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   CELL DIVISION PROTEIN FTSZ.
GN   FTSZ.
OS   Azotobacter vinelandii.
OC   Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae;
OC   Azotobacter.
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=DJ116;
RX   MEDLINE; 98267010.
RA   LU C., STRICKER J., ERICKSON H.P.;
RT   "FtsZ from Escherichia coli, Azotobacter vinelandii, and Thermotoga
RT   maritima -- quantitation, GTP hydrolysis, and assembly.";
RL   Cell Motil. Cytoskeleton 40:71-86(1998).
CC   -!- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.
CC       ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF
CC       THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,
CC       AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPTATION TO
CC       BEGIN. BINDS TO AND HYDROLYZES GTP.
CC   -!- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSEMBLE AT THE INNER SURFACE
CC       OF THE CYTOPLASMIC MEMBRANE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U65939; AAC24603.1; -.
DR   HSSP; Q57816; 1FSZ.
DR   PROSITE; PS01134; FTSZ_1; 1.
DR   PROSITE; PS01135; FTSZ_2; 1.
KW   Cell division; Septation; GTP-binding.
FT   NP_BIND     104    112       GTP (POTENTIAL).
SQ   SEQUENCE   394 AA;  41153 MW;  4E887134 CRC32;

  Query Match           71.0%;  Score 44;  DB 1;  Length 394;
  Best Local Similarity  77.8%;  Pred. No. 2.41e+01;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    121 AKGLGILTV 129
          | |:|||||
Qy      2 AAGIGILTV 10


RESULT   13
ID   GLMU_ECOLI     STANDARD;      PRT;   456 AA.
AC   P17114; P76746;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE PYROPHOSPHORYLASE (EC 2.7.7.23) (N-
DE   ACETYLGLUCOSAMINE-1-PHOSPHATE URIDYLTRANSFERASE).
GN   GLMU.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85121806.
RA   WALKER J.E., GAY N.J., SARASTE M., EBERLE A.N.;
RT   "DNA sequence around the Escherichia coli unc operon. Completion of
RT   the sequence of a 17 kilobase segment containing asnA, oriC, unc,
RT   glmS and phoS.";
RL   Biochem. J. 224:799-815(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 93315143.
RA   BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT   "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT   genome: organizational symmetry around the origin of replication.";
RL   Genomics 16:551-561(1993).
RN   [3]
RP   REVISIONS.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   IDENTIFICATION.
RX   MEDLINE; 94012475.
RA   MENGIN-LECREULX D., VAN HEIJENOORT J.;
RT   "Identification of the glmU gene encoding N-acetylglucosamine-1-
RT   phosphate uridyltransferase in Escherichia coli.";
RL   J. Bacteriol. 175:6150-6157(1993).
CC   -!- FUNCTION: BIFUNCTIONAL ENZYME RESPONSIBLE FOR THE ACETYLATION OF
CC       GLC-N-1-P TO GIVE GLCNAC-1-P AND THE SYNTHESIS OF UDP-GLCNAC.
CC   -!- CATALYTIC ACTIVITY: UTP + N-ACETYL-ALPHA-D-GLUCOSAMINE
CC       1-PHOSPHATE = PYROPHOSPHATE + UDP-N-ACETYL-D-GLUCOSAMINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN AND LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF
CC       ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT THAT CREATES TWO ORFS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X01631; CAA25784.1; -.
DR   EMBL; L10328; AAA62082.1; ALT_FRAME.
DR   EMBL; L10328; AAA62081.1; ALT_FRAME.
DR   EMBL; AE000450; AAC76753.1; -.
DR   ECOGENE; EG11198; GLMU.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; 1.
```

```
DR   PFAM; PF00132; hexapep; 3.
DR   PFAM; PF00483; NTP_transferase; 1.
KW   Peptidoglycan synthesis; Cell wall; Transferase;
KW   Nucleotidyltransferase; Repeat; Multifunctional enzyme.
FT   CONFLICT    186    187       KL -> NV (IN REF. 1).
SQ   SEQUENCE   456 AA;  49190 MW;  B9E65439 CRC32;

  Query Match           71.0%;  Score 44;  DB 1;  Length 456;
  Best Local Similarity  75.0%;  Pred. No. 2.41e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    124 GGIGLLTV 131
          :|||:|||
Qy      3 AGIGILTV 10


RESULT   14
ID   FLID_ECOLI     STANDARD;      PRT;   467 AA.
AC   P24216;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   FLAGELLAR HOOK-ASSOCIATED PROTEIN 2 (HAP2) (FILAMENT CAP PROTEIN).
GN   FLID OR FLBC OR FLAV.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JA11;
RX   MEDLINE; 92407478.
RA   KAWAGISHI I., MUELLER V., WILLIAMS A.W., IRIKURA V.M., MACNAB R.M.;
RT   "Subdivision of flagellar region III of the Escherichia coli and
RT   Salmonella typhimurium chromosomes and identification of two
RT   additional flagellar genes.";
RL   J. Gen. Microbiol. 138:1051-1065(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE OF 1-152 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE; 89281489.
RA   HANAFUSA T., SAKAI A., TOMINAGA A., ENOMOTO M.;
RT   "Isolation and characterization of Escherichia coli hag operator
RT   mutants whose hag48 expression has become repressible by a Salmonella
RT   H1 repressor.";
RL   Mol. Gen. Genet. 216:44-50(1989).
RN   [4]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE; 83238225.
RA   SZEKELY E., SIMON M.;
RT   "DNA sequence adjacent to flagellar genes and evolution of flagellar-
RT   phase variation.";
RL   J. Bacteriol. 155:74-81(1983).
CC   -!- FUNCTION: CAPPING PROTEIN FOR THE FLAGELLA; FORMS THE DISTAL END
CC       OF THE FLAGELLA.
CC   -!- SIMILARITY: TO OTHER FILAMENT CAP PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M85240; AAA23790.1; -.
DR   EMBL; AE000285; AAC74991.1; -.
DR   EMBL; X17440; CAA35487.1; -.
DR   EMBL; J01607; AAA92490.1; -.
DR   PIR; PV0005; PV0005.
DR   ECOGENE; EG10841; FLID.
KW   Flagella.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CONFLICT    113    113       T -> R (IN REF. 3).
SQ   SEQUENCE   467 AA;  48270 MW;  14800A2E CRC32;

  Query Match           71.0%;  Score 44;  DB 1;  Length 467;
  Best Local Similarity  50.0%;  Pred. No. 2.41e+01;
  Matches     5;  Conservative    3;  Mismatches   2;  Indels   0;  Gaps   0;

Db    368 DASGVGALIV 377
          :|:|:| | |
Qy      1 EAAGIGILTV 10


RESULT   15
ID   XYND_PAEPO     STANDARD;      PRT;   635 AA.
AC   P45796;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ENDO-1,4-BETA-XYLANASE D PRECURSOR (EC 3.2.1.8) (XYLANASE D)
DE   (1,4-BETA-D-XYLAN XYLANOHYDROLASE D).
GN   XYND.
OS   Paenibacillus polymyxa (Bacillus polymyxa).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Paenibacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 842;
RX   MEDLINE; 92041687.
RA   GOSALBES M.J., PEREZ-GONZALEZ J.A., GONZALEZ R., NAVARRO A.;
RT   "Two beta-glycanase genes are clustered in Bacillus polymyxa:
RT   molecular cloning, expression, and sequence analysis of genes
RT   encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
RL   J. Bacteriol. 173:7705-7710(1991).
CC   -!- FUNCTION: SHOWS XYLANASE ACTIVITY AS WELL AS ALPHA-L-
CC       ARABINOFURANOSIDASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC       LINKAGES IN XYLANS.
CC   -!- PATHWAY: XYLAN DEGRADATION.
CC   -!- SIMILARITY: BELONGS TO FAMILY 43 OF GLYCOSYL HYDROLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X57094; CAA40378.1; -.
KW   Xylan degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    635       ENDO-1,4-BETA-XYLANASE D.
SQ   SEQUENCE   635 AA;  67914 MW;  078AAB82 CRC32;

  Query Match           71.0%;  Score 44;  DB 1;  Length 635;
  Best Local Similarity  75.0%;  Pred. No. 2.41e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    149 GAGIGVLT 156
          :||||:||
Qy      2 AAGIGILT 9


Search completed: Fri May  5 22:09:57 2000
```

Job time : 39 secs.

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:08:01 2000; MasPar time 4.93 Seconds
95.764 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-17
Description: (1-10) from US09267439.pep
Perfect Score: 62
Sequence: 1 EAAGIGILTV 10

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.809; Variance 26.769; scale 0.852

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	62	100.0	118	2	A55253	melanoma antigen MART	6.18e-03
2	48	77.4	333	1	DEALXE	acetoin[2,6-dichlorop	9.75e+00
3	47	75.8	101	2	A33351	H+-transporting ATP s	1.57e+01
4	47	75.8	110	1	VHBPDL	major capsid protein	1.57e+01
5	47	75.8	165	2	C70959	hypothetical protein	1.57e+01
6	47	75.8	250	2	A69843	hypothetical protein	1.57e+01
7	47	75.8	479	1	VGBEPB	glycoprotein gIII pre	1.57e+01
8	47	75.8	753	1	A42863	5-methyltetrahydropte	1.57e+01
9	47	75.8	773	2	T00502	protein kinase homolo	1.57e+01
10	46	74.2	231	1	S48276	YSA1 protein - yeast	2.52e+01
11	46	74.2	271	2	S38115	hypothetical protein	2.52e+01
12	45	72.6	201	2	A71124	hypothetical protein	4.01e+01
13	45	72.6	339	2	S62369	methylcobalamin--coen	4.01e+01
14	45	72.6	420	2	S59131	Kan-1 protein - rat	4.01e+01
15	45	72.6	744	2	A43353	ascites sialoglycopro	4.01e+01
16	44	71.0	98	2	D72106	hypothetical protein	6.33e+01
17	44	71.0	456	2	C65176	glmU protein - Escher	6.33e+01
18	44	71.0	468	2	A64956	flagellar hook-associ	6.33e+01
19	44	71.0	620	2	H69382	ABC transporter, ATP-	6.33e+01
20	44	71.0	635	2	S19011	endo-1,4-beta-xylanas	6.33e+01
21	43	69.4	123	2	A71312	probable anti-sigma F	9.91e+01
22	43	69.4	132	2	S01903	H+-transporting ATP s	9.91e+01
23	43	69.4	207	2	H69694	ribosomal protein L4	9.91e+01
24	43	69.4	231	2	G75193	hypothetical protein	9.91e+01
25	43	69.4	345	1	KZEC	alkaline phosphatase	9.91e+01
26	43	69.4	402	2	D70602	probable arginine dei	9.91e+01
27	43	69.4	461	2	D70073	metabolite transport	9.91e+01
28	43	69.4	461	2	T03561	hypothetical protein	9.91e+01
29	43	69.4	493	1	ACMSE	nicotinic acetylcholi	9.91e+01
30	43	69.4	509	2	H70597	probable membrane pro	9.91e+01
31	43	69.4	580	2	B70868	probable transferase	9.91e+01
32	43	69.4	590	2	A45772	nitrate-inducible nit	9.91e+01
33	43	69.4	593	1	A69655	two-component sensor	9.91e+01
34	43	69.4	611	2	JT0592	hypothetical protein	9.91e+01
35	43	69.4	667	2	F70682	probable membrane pro	9.91e+01
36	43	69.4	675	2	S53832	NADH dehydrogenase (u	9.91e+01
37	43	69.4	729	2	T06127	probable sugar transp	9.91e+01
38	43	69.4	746	2	T06017	subtilisin homolog -	9.91e+01
39	43	69.4	1217	1	EGMSMG	epidermal growth fact	9.91e+01
40	43	69.4	1325	2	A64905	ydeK protein - Escher	9.91e+01
41	43	69.4	1436	2	S67655	probable membrane pro	9.91e+01
42	43	69.4	1616	2	G70668	polyketide synthase p	9.91e+01
43	42	67.7	218	1	B41316	flagellin B1 precurso	1.54e+02
44	42	67.7	251	2	S56416	hypothetical transcri	1.54e+02
45	42	67.7	417	2	D70321	sulfide dehydrogenase	1.54e+02

ALIGNMENTS

```
RESULT    1
ENTRY             A55253     #type complete
TITLE             melanoma antigen MART-1 - human
ALTERNATE_NAMES   melan-A protein
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
                    10-Sep-1997
ACCESSIONS        A55253; I38506
REFERENCE         A55253
   #authors       Kawakami, Y.; Eliyahu, S.; Delgado, C.H.; Robbins, P.F.;
                    Rivoltini, L.; Topalian, S.L.; Miki, T.; Rosenberg, S.A.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1994) 91:3515-3519
   #title         Cloning of the gene coding for a shared human melanoma
                    antigen recognized by autologous T cells infiltrating into
                    tumor.
   #cross-references MUID:94224770
   #accession     A55253
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-118 ##label KAW
      ##cross-references GB:U06452; NID:g476131; PID:g476132
REFERENCE         I38506
   #authors       Coulie, P.G.; Brichard, V.; Van Pel, A.; Wolfel, T.;
                    Schneider, J.; Traversari, C.; Mattei, S.; De Plaen, E.;
                    Lurquin, C.; Szikora, J.P.; Renauld, J.; Boon, T.
   #journal       J. Exp. Med. (1994) 180:35-42
   #title         A new gene coding for a differentiation antigen recognized by
                    autologous cytolytic T lymphocytes on HLA-A2 melanomas [see
                    comments].
   #cross-references MUID:94275389
   #accession     I38506
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-118 ##label RES
      ##cross-references EMBL:U06654; NID:g517022; PID:g517023
GENETICS
   #gene          GDB:MLANA
      ##cross-references GDB:358979
   #map_position 17q21-17q24
SUMMARY           #length 118  #molecular-weight 13157  #checksum 3535

  Query Match           100.0%;  Score 62;  DB 2;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 6.18e-03;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      26 EAAGIGILTV 35
           ||||||||||
```

```
Qy      1 EAAGIGILTV 10


RESULT    2
ENTRY             DEALXE     #type complete
TITLE             acetoin[2,6-dichlorophenolindophenol] oxidoreductase (EC
                    1.-.-.-) alpha chain - Alcaligenes eutrophus (strain H16)
ORGANISM          #formal_name Alcaligenes eutrophus
DATE              31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
                    11-Jun-1999
ACCESSIONS        B42462
REFERENCE         A42462
   #authors       Priefert, H.; Hein, S.; Krueger, N.; Zeh, K.; Schmidt, B.;
                    Steinbuechel, A.
   #journal       J. Bacteriol. (1991) 173:4056-4071
   #title         Identification and molecular characterization of the
                    Alcaligenes eutrophus H16 aco operon genes involved in
                    acetoin catabolism.
   #cross-references MUID:91286190
   #accession     B42462
      ##molecule_type DNA
      ##residues       1-333 ##label PRI
      ##cross-references GB:M66060; NID:g141892; PIDN:AAA21948.1; PID:g141894
COMMENT     This is a component of the enzyme complex that catalyzes 2,
              6-dichlorophenolindophenol-dependent cleavage of acetoin into
              acetate and acetaldehyde. The functional enzyme is a tetramer of
              two alpha and two beta chains.
CLASSIFICATION    #superfamily pyruvate dehydrogenase (lipoamide) alpha chain;
                    thiamine pyrophosphate-binding domain homology
KEYWORDS          heterotetramer; oxidoreductase
FEATURE
   145-194             #domain thiamine pyrophosphate-binding domain homology
                         #label TPB
SUMMARY           #length 333  #molecular-weight 35375  #checksum 2647

  Query Match           77.4%;  Score 48;  DB 1;  Length 333;
  Best Local Similarity  87.5%;  Pred. No. 9.75e+00;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     56 EAAGVGIL 63
          ||||:|||
Qy      1 EAAGIGIL 8


RESULT    3
ENTRY             A33351     #type complete
TITLE             H+-transporting ATP synthase (EC 3.6.1.34) proteolipid chain
                    - Sulfolobus acidocaldarius
ORGANISM          #formal_name Sulfolobus acidocaldarius
DATE              20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change
                    22-Jun-1999
ACCESSIONS        A33351
REFERENCE         A33351
   #authors       Denda, K.; Konishi, J.; Oshima, T.; Date, T.; Yoshida, M.
   #journal       J. Biol. Chem. (1989) 264:7119-7121
   #title         A gene encoding the proteolipid subunit of Sulfolobus
                    acidocaldarius ATPase complex.
   #cross-references MUID:89214142
   #accession     A33351
      ##status        preliminary
      ##molecule_type DNA
      ##residues       1-101 ##label DEN
      ##cross-references GB:J04740; NID:g152922; PIDN:AAA72703.1; PID:g152925
CLASSIFICATION    #superfamily H+-transporting ATP synthase lipid-binding
                    protein
KEYWORDS          hydrolase
SUMMARY           #length 101  #molecular-weight 10362  #checksum 4300

  Query Match           75.8%;  Score 47;  DB 2;  Length 101;
  Best Local Similarity  87.5%;  Pred. No. 1.57e+01;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 AAGIGVLT 66
          |||||:||
Qy      2 AAGIGILT 9


RESULT    4
ENTRY             VHBPDL     #type complete
TITLE             major capsid protein D - phage lambda
ALTERNATE_NAMES   head protein D
ORGANISM          #formal_name phage lambda
DATE              13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change
                    23-Jul-1999
ACCESSIONS        G04333; C43013; A04334; A23206
REFERENCE         A94614
   #authors       Daniels, D.
   #submission    submitted to the Nucleic Acid Sequence Database, September
                    1982
   #accession     G04333
      ##molecule_type DNA
      ##residues       1-110 ##label DAN
REFERENCE         A92891
   #authors       Sanger, F.; Coulson, A.R.; Hong, G.F.; Hill, D.F.; Petersen,
                    G.B.
   #journal       J. Mol. Biol. (1982) 162:729-773
   #title         Nucleotide sequence of bacteriophage lambda DNA.
   #cross-references MUID:83189071
   #accession     C43013
      ##molecule_type DNA
      ##residues       1-110 ##label SAN
      ##cross-references GB:J02459; GB:M17233; GB:M24325; GB:V00636;
                          GB:X00906; NID:g215104; PIDN:AAA96539.1; PID:g215111
REFERENCE         A23206
   #authors       Witkiewicz, H.; Schweiger, M.
   #journal       EMBO J. (1982) 1:1559-1564
   #cross-references MUID:84207913
   #contents      annotation; physicochemical properties
COMMENT     Gene D protein is a major component of the phage head and serves to
              stabilize the head during DNA packaging. There are approximately
              420 copies of protein D per mature phage.
GENETICS
   #gene          D
   #map_position 11.85-12.53
CLASSIFICATION    #superfamily phage lambda major capsid protein D
KEYWORDS          DNA packaging
SUMMARY           #length 110  #molecular-weight 11572  #checksum 3863

  Query Match           75.8%;  Score 47;  DB 1;  Length 110;
  Best Local Similarity  50.0%;  Pred. No. 1.57e+01;
  Matches     5;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db     55 DGAAVGILAV 64
          ::|::|||:|
Qy      1 EAAGIGILTV 10


RESULT    5
ENTRY             C70959     #type complete
TITLE             hypothetical protein Rv1382 - Mycobacterium tuberculosis
                    (strain H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    26-Aug-1999
ACCESSIONS        C70959
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
```

```
   #journal      Nature (1998) 393:537-544
   #title        Deciphering the biology of Mycobacterium tuberculosis from
                   the complete genome sequence.
   #cross-references MUID:98295987
   #accession    C70959
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-165 ##label COL
      ##cross-references GB:Z81011; GB:AL123456; NID:g3242274; PID:e275153;
                          PID:g1621264
      ##experimental_source strain H37Rv
GENETICS
   #gene         Rv1382
CLASSIFICATION   #superfamily Mycobacterium tuberculosis hypothetical protein
                   Rv1382
SUMMARY          #length 165  #molecular-weight 18189  #checksum 5780

  Query Match           75.8%;  Score 47;  DB 2;  Length 165;
  Best Local Similarity  75.0%;  Pred. No. 1.57e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    128 AGIGILAI 135
          ||||||::
Qy      3 AGIGILTV 10


RESULT    6
ENTRY            A69843     #type complete
TITLE            hypothetical protein yjbA - Bacillus subtilis
ORGANISM         #formal_name Bacillus subtilis
DATE             05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                   24-Sep-1998
ACCESSIONS       A69843
REFERENCE        A69580
   #authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                   Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                   Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                   A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                   Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                   Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                   Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                   Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                   Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                   M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                   S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                   Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                   C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                   Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                   Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                   Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                   Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                   Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                   Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                   M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                   M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                   V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                   A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                   Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                   Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                   Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                   Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                   B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                   Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                   Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                   Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                   Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                   Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                   K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                   Yoshikawa, H.; Danchin, A.
   #journal      Nature (1997) 390:249-256
   #title        The complete genome sequence of the Gram-positive bacterium
                   Bacillus subtilis.
   #cross-references MUID:98044033
   #accession    A69843
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-250 ##label KUN
      ##cross-references GB:Z99110; GB:AL009126; NID:g2633472; PID:e1183161;
                          PID:g2633495
      ##experimental_source strain 168
GENETICS
   #gene         yjbA
SUMMARY          #length 250  #molecular-weight 30119  #checksum 5271

  Query Match           75.8%;  Score 47;  DB 2;  Length 250;
  Best Local Similarity  66.7%;  Pred. No. 1.57e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     89 TDGIGILAV 97
          : |||||:|
Qy      2 AAGIGILTV 10


RESULT    7
ENTRY            VGBEPB     #type complete
TITLE            glycoprotein gIII precursor - suid herpesvirus 1
ORGANISM         #formal_name suid herpesvirus 1
DATE             30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
                   16-Jul-1999
ACCESSIONS       A26097
REFERENCE        A26097
   #authors      Robbins, A.K.; Watson, R.J.; Whealy, M.E.; Hays, W.W.;
                   Enquist, L.W.
   #journal      J. Virol. (1986) 58:339-347
   #title        Characterization of a Pseudorabies virus glycoprotein gene
                   with homology to herpes simplex virus type 1 and type 2
                   glycoprotein C.
   #cross-references MUID:86200375
   #accession    A26097
      ##molecule_type DNA
      ##residues      1-479 ##label ROB
      ##cross-references GB:M12778; NID:g334049; PIDN:AAA47464.1; PID:g334050
      ##experimental_source strain Becker
CLASSIFICATION   #superfamily herpesvirus glycoprotein F
KEYWORDS         glycoprotein
FEATURE
   1-22              #domain signal sequence #status predicted #label SIG\
   23-479            #product glycoprotein gIII #status predicted #label GPG\
   40,84,169,192,220,
   228,285,302       #binding_site carbohydrate (Asn) (covalent) #status
                       predicted
SUMMARY          #length 479  #molecular-weight 51206  #checksum 1630

  Query Match           75.8%;  Score 47;  DB 1;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 1.57e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      3 AGIGILTV 10


RESULT    8
ENTRY            A42863     #type complete
TITLE            5-methyltetrahydropteroyltriglutamate--homocysteine
                   S-methyltransferase (EC 2.1.1.14) - Escherichia coli
                   (strain K-12)
ALTERNATE_NAMES  cobalamin-independent methionine synthase;
                   tetrahydropteroylglutamate methyltransferase
ORGANISM         #formal_name Escherichia coli
DATE             17-Feb-1994 #sequence_revision 10-Oct-1997 #text_change
                   11-Jun-1999
ACCESSIONS       F65187; A42863; S30719; I79560
REFERENCE        A64720
```

```
   #authors      Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                   Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                   Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                   Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                   Y.
   #journal      Science (1997) 277:1453-1462
   #title        The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession    F65187
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-753 ##label BLAT
      ##cross-references GB:AE000458; GB:U00096; NID:g2367299;
                        PIDN:AAC76832.1; PID:g2367304; UWGP:b3829
      ##experimental_source strain K-12, substrain MG1655
REFERENCE        A42863
   #authors      Gonzalez, J.C.; Banerjee, R.V.; Huang, S.; Sumner, J.S.;
                   Matthews, R.G.
   #journal      Biochemistry (1992) 31:6045-6056
   #title        Comparison of cobalamin-independent and cobalamin-dependent
                   methionine synthases from Escherichia coli: two solutions
                   to the same chemical problem.
   #cross-references MUID:92329421
   #accession    A42863
      ##molecule_type DNA
      ##residues      1-362,'V',364-753 ##label GON
      ##cross-references GB:M87625; NID:g304870; PIDN:AAA23544.1; PID:g145474
      ##experimental_source strain DH5alphaF'
      ##note          sequence extracted from NCBI backbone (NCBIP:109176)
REFERENCE        S30660
   #authors      Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
   #journal      Science (1992) 257:771-778
   #title        Analysis of the Escherichia coli genome: DNA sequence of the
                   region from 84.5 to 86.5 minutes.
   #cross-references MUID:92358234
   #accession    S30719
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-580,'CVMKWPIWKPLELASSRLTNRR',603-604,'Q',606-658,'R',
                        660-753 ##label DAN
      ##cross-references EMBL:M87049
      ##note          the nucleotide sequence was submitted to the EMBL Data
                        Library, November 1992
      ##note          this sequence has been corrected in reference A64720
REFERENCE        I59156
   #authors      Maxon, M.E.; Redfield, B.; Cai, X.
   #journal      Proc. Natl. Acad. Sci. U.S.A. (1989) 86:85-89
   #title        Regulation of methionine synthesis in Escherichia coli:
                   Effect of the MetR protein on the expression of the metE
                   and metR genes.
   #cross-references MUID:89098936
   #accession    I79560
      ##status        preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-22 ##label RES
      ##cross-references GB:J04155; NID:g146825; PIDN:AAA24160.1; PID:g146827
GENETICS
   #gene         metE
CLASSIFICATION   #superfamily cobalamin-independent methionine synthase
KEYWORDS         methionine biosynthesis; methylated amino acid;
                   methyltransferase

FEATURE
   726               #active_site Cys (methylcysteine intermediate) #status
                        experimental
SUMMARY          #length 753  #molecular-weight 84673  #checksum 6435

  Query Match           75.8%;  Score 47;  DB 1;  Length 753;
  Best Local Similarity  70.0%;  Pred. No. 1.57e+01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    589 EAAGIGIIQI 598
          |||||||: :
Qy      1 EAAGIGILTV 10
```

```
RESULT    9
ENTRY            T00502     #type complete
TITLE            protein kinase homolog T20D16.7 - Arabidopsis thaliana
ORGANISM         #formal_name Arabidopsis thaliana #common_name mouse-ear
                   cress
DATE             01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
                   23-Apr-1999
ACCESSIONS       T00502
REFERENCE        Z14159
   #authors      Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
                   Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
                   Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
                   J.C.
   #submission   submitted to the EMBL Data Library, November 1997
   #description  Arabidopsis thaliana chromosome II BAC T20D16 genomic
                   sequence.
   #accession    T00502
      ##status        translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues      1-773 ##label ROU
      ##cross-references EMBL:AC002391; NID:g2642427; PID:g2642433
      ##experimental_source cultivar Columbia
GENETICS
   #map_position 2
   #introns      545/1
   #note         T20D16.7
SUMMARY          #length 773  #molecular-weight 84148  #checksum 123

  Query Match           75.8%;  Score 47;  DB 2;  Length 773;
  Best Local Similarity  60.0%;  Pred. No. 1.57e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    341 DIAGIGILAL 350
          : ||||||::
Qy      1 EAAGIGILTV 10


RESULT   10
ENTRY            S48276     #type complete
TITLE            YSA1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES  protein YBR0907; protein YBR111c
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
                   10-Sep-1999
ACCESSIONS       S48276; S45979; S25364; S44691
REFERENCE        S48255
   #authors      Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann,
                   H.
   #journal      Yeast (1994) 10:1363-1381
   #title        Analysis of a 70 kb region on the right arm of yeast
                   chromosome II.
   #cross-references MUID:95208357
   #accession    S48276
      ##status        nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues      1-231 ##label MAN
      ##cross-references EMBL:X78993; NID:g476045; PID:g476067
REFERENCE        S45927
   #authors      Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
   #submission   submitted to the Protein Sequence Database, August 1994
   #accession    S45979
      ##molecule_type DNA
      ##residues      1-231 ##label FE2
      ##cross-references EMBL:Z35980; NID:g536465; PID:g536466; MIPS:YBR111c
REFERENCE        S25364
   #authors      Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann,
                   H.
   #journal      Yeast (1992) 8:397-408
   #title        Molecular analysis of yeast chromosome II between CMD1 and
                   LYS2: the excision repair gene RAD16 located in this region
                   belongs to a novel group of double-finger proteins.
```

```
   #cross-references MUID:92327848
   #accession    S25364
      ##molecule_type DNA
      ##residues      1-47 ##label MAW
      ##cross-references EMBL:X66247; NID:g3548; PID:g3549
GENETICS
   #gene         SGD:YSA1
      ##cross-references SGD:S0000315; MIPS:YBR111c
   #map_position 2R
CLASSIFICATION   #superfamily yffH protein; mutT domain homology
FEATURE
   111-145           #domain mutT domain homology #label MUTT
SUMMARY          #length 231  #molecular-weight 26087  #checksum 4809

  Query Match           74.2%;  Score 46;  DB 1;  Length 231;
  Best Local Similarity  85.7%;  Pred. No. 2.52e+01;
  Matches     6;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     79 GIGILTI 85
          |||||||:
Qy      4 GIGILTV 10


RESULT   11
ENTRY            S38115     #type complete
TITLE            hypothetical protein YKR043c - yeast (Saccharomyces
                   cerevisiae)
ORGANISM         #formal_name Saccharomyces cerevisiae
DATE             03-May-1994 #sequence_revision 03-May-1994 #text_change
                   14-Nov-1997
ACCESSIONS       S38115
REFERENCE        S38097
   #authors      Urrestarazu, L.A.; Jauniaux, J.C.
   #submission   submitted to the Protein Sequence Database, March 1994
   #accession    S38115
      ##molecule_type DNA
      ##residues      1-271 ##label URR
      ##cross-references EMBL:Z28268; NID:g486490; PID:g486491; MIPS:YKR043c
      ##experimental_source strain S288C
GENETICS
   #map_position 11R
SUMMARY          #length 271  #molecular-weight 31022  #checksum 8533

  Query Match           74.2%;  Score 46;  DB 2;  Length 271;
  Best Local Similarity  55.6%;  Pred. No. 2.52e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    233 DAGGIGVLS 241
          :|:|||:|:
Qy      1 EAAGIGILT 9


RESULT   12
ENTRY            A71124     #type complete
TITLE            hypothetical protein PH0760 - Pyrococcus horikoshii
ORGANISM         #formal_name Pyrococcus horikoshii
DATE             14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
                   26-Aug-1999
ACCESSIONS       A71124
REFERENCE        A71000
   #authors      Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
                   Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                   Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                   Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                   Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                   A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                   Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
   #journal      DNA Res. (1998) 5:55-76
   #title        Complete sequence and gene organization of the genome of a
                   hyper-thermophilic archaebacterium, Pyrococcus horikoshii
                   OT3.
   #cross-references MUID:98344137
   #accession    A71124
      ##status       preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-201 ##label KAW
      ##cross-references GB:AP000003; NID:g3236130; PID:d1030794; PID:g3257168
      ##experimental_source strain OT3
      ##note          this accession replaces an interim accession for a
                       sequence replaced by GenBank
GENETICS
   #gene         PH0760
CLASSIFICATION   #superfamily conserved hypothetical protein MJ1677
SUMMARY          #length 201  #molecular-weight 21592  #checksum 3142

  Query Match           72.6%;  Score 45;  DB 2;  Length 201;
  Best Local Similarity  60.0%;  Pred. No. 4.01e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    137 ESPGIVILTI 146
          |::|| |||:
Qy      1 EAAGIGILTV 10


RESULT   13
ENTRY            S62369     #type complete
TITLE            methylcobalamin--coenzyme M methyltransferase II -
                   Methanosarcina barkeri
ORGANISM         #formal_name Methanosarcina barkeri
DATE             19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
                   17-Mar-1999
ACCESSIONS       S62369
REFERENCE        S62368
   #authors      Harms, U.; Thauer, R.K.
   #journal      Eur. J. Biochem. (1996) 235:653-659
   #title        Methylcobalamin:coenzyme M methyltransferase isoenzymes MtaA
                   and MtbA from Methanosarcina barkeri: cloning, sequencing
                   and differential transcription of the encoding genes, and
                   functional overexpression of the mtaA gene in Escherichia
                   coli.
   #cross-references MUID:96184544
   #accession    S62369
      ##status       preliminary; nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues      1-339 ##label HAR
      ##cross-references EMBL:X91894; NID:g1107727; PID:e204100; PID:g1107728
SUMMARY          #length 339  #molecular-weight 36761  #checksum 6431

  Query Match           72.6%;  Score 45;  DB 2;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 4.01e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      3 AGIGILTV 10


RESULT   14
ENTRY            S59131     #type complete
TITLE            Kan-1 protein - rat
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
                   07-May-1999
ACCESSIONS       S59131
REFERENCE        S59131
   #authors      Furutani, M.; Arii, S.; Higashitsuji, H.; Mise, M.; Fukumoto,
                   M.; Takano, S.; Nakayama, H.; Imamura, M.; Fujita, J.
   #journal      Biochem. J. (1995) 311:203-208
   #title        Reduced expression of kan-1 (encoding putative bile
                   acid-CoA-amino acid N-acyltransferase) mRNA in livers of
                   rats after partial hepatectomy and during sepsis.
   #cross-references MUID:96003917
   #accession    S59131
      ##status       preliminary
      ##molecule_type mRNA
```

```
     ##residues      1-420 ##label FUR
     ##cross-references EMBL:D43964; NID:g604901; PID:d1008487; PID:g604902
SUMMARY          #length 420  #molecular-weight 46496  #checksum 4868

  Query Match           72.6%;  Score 45;  DB 2;  Length 420;
  Best Local Similarity  55.6%;  Pred. No. 4.01e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    226 GPGVGILSV 234
          ::|:|||:|
Qy      2 AAGIGILTV 10


RESULT   15
ENTRY            A43353     #type fragment
TITLE            ascites sialoglycoprotein-2 - rat (fragment)
ORGANISM         #formal_name Rattus norvegicus #common_name Norway rat
DATE             10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change
                   16-Jul-1999
ACCESSIONS       A43353
REFERENCE        A43353
   #authors      Sheng, Z.; Wu, K.; Carraway, K.L.; Fregien, N.
   #journal      J. Biol. Chem. (1992) 267:16341-16346
   #title        Molecular cloning of the transmembrane component of the 13762
                   mammary adenocarcinoma sialomucin complex. A new member of
                   the epidermal growth factor superfamily.
   #cross-references MUID:92355597
   #accession    A43353
      ##status         preliminary
      ##molecule_type mRNA; protein
      ##residues       1-744 ##label SHE
      ##experimental_source mammary adenocarcinoma
      ##note           sequence extracted from NCBI backbone (NCBIN:110690,
                        NCBIP:110691)
CLASSIFICATION   #superfamily EGF homology
KEYWORDS         glycoprotein
FEATURE
   655-694            #domain EGF homology #label EGF
SUMMARY          #length 744  #checksum 2462

  Query Match           72.6%;  Score 45;  DB 2;  Length 744;
  Best Local Similarity  60.0%;  Pred. No. 4.01e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    201 ETNGIGLLGV 210
          |: |||:| |
Qy      1 EAAGIGILTV 10


Search completed: Fri May  5 22:09:02 2000
Job time : 61 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:18:37 2000; MasPar time 7.35 Seconds
94.346 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-18
Description: (1-10) from US09267439.pep
Perfect Score: 63
Sequence: 1 AAGIGILTVI 10

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 22.816; Variance 28.079; scale 0.813

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	56	88.9	766	10	O23161	RECEPTOR KINASE-LIKE P	4.47e-01
2	52	82.5	478	14	Q87090	GLYCOPROTEIN GIII.	3.29e+00
3	52	82.5	479	14	Q87089	GLYCOPROTEIN GIII.	3.29e+00
4	52	82.5	479	14	Q87091	GLYCOPROTEIN GIII.	3.29e+00
5	50	79.4	509	2	O05457	HYPOTHETICAL 53.2 KD P	8.61e+00
6	50	79.4	773	10	O22178	PUTATIVE RECEPTOR PROT	8.61e+00
7	49	77.8	295	5	Q09957	HYPOTHETICAL 32.3 KD P	1.38e+01
8	48	76.2	344	2	Q9X4R2	HEAT SHOCK TRANSCRIPTI	2.20e+01
9	48	76.2	361	2	O31536	YETI PROTEIN.	2.20e+01
10	48	76.2	521	10	O22848	MEMBRANE TRANSPORTER D	2.20e+01
11	48	76.2	667	2	P71749	HYPOTHETICAL 68.3 KD P	2.20e+01
12	48	76.2	677	14	Q89853	VIRION SPIKE GLYCOPROT	2.20e+01
13	48	76.2	677	14	Q66799	VIRION SPIKE GLYCOPROT	2.20e+01
14	48	76.2	808	10	Q9ZT37	PUTATIVE GLUTAMATE REC	2.20e+01
15	47	74.6	91	2	Q49684	B1496_C2_163.	3.48e+01
16	47	74.6	165	2	P71810	HYPOTHETICAL 18.2 KD P	3.48e+01
17	47	74.6	226	5	O17993	R08H2.11 PROTEIN.	3.48e+01
18	47	74.6	250	2	O31597	YJBA PROTEIN.	3.48e+01
19	47	74.6	251	2	Q9X3X6	MOLYBDOPTERIN BIOSYNTH	3.48e+01
20	47	74.6	370	8	O48172	CYTOCHROME B.	3.48e+01
21	47	74.6	503	2	O26074	PROTEIN-EXPORT MEMBRAN	3.48e+01
22	47	74.6	526	2	Q9ZJ66	PROTEIN-EXPORT MEMBRAN	3.48e+01
23	46	73.0	29	2	Q08008	EXPORT ELEMENT BL13 (F	5.46e+01
24	46	73.0	86	10	P93048	GAG1AT MRNA.	5.46e+01
25	46	73.0	230	5	O16265	F40A3.2 PROTEIN.	5.46e+01
26	46	73.0	271	2	Q9X130	CONSERVED HYPOTHETICAL	5.46e+01
27	46	73.0	287	1	Q9YFL7	287AA LONG HYPOTHETICA	5.46e+01
28	46	73.0	291	2	Q9ZFP3	EFMB.	5.46e+01
29	46	73.0	637	10	Q03678	EMBRYO GLOBULIN.	5.46e+01
30	46	73.0	779	2	O32231	YVAJ PROTEIN.	5.46e+01
31	46	73.0	848	5	O18139	T26H2.7 PROTEIN.	5.46e+01
32	45	71.4	98	2	Q9Z8X3	HYPOTHETICAL 10.3 KD P	8.51e+01
33	45	71.4	217	2	O67058	HYPOTHETICAL 24.8 KD P	8.51e+01
34	45	71.4	336	5	O17857	F28H7.11 PROTEIN.	8.51e+01
35	45	71.4	339	1	Q48950	METHYLCOBALAMIN: COENZ	8.51e+01
36	45	71.4	339	1	O30640	METHYLCOBAMIDE:COM MET	8.51e+01
37	45	71.4	339	1	Q48928	METHYLCOBAMIDE:COM MET	8.51e+01
38	45	71.4	407	2	Q9X5P8	CYTOCHROME P450 HYDROX	8.51e+01
39	45	71.4	420	11	Q63276	KAN-1.	8.51e+01
40	45	71.4	420	11	O08833	BILE ACID COA: AMINO A	8.51e+01
41	45	71.4	460	2	O84908	GLYCOSYLTRANSFERASE WB	8.51e+01
42	45	71.4	500	2	O86074	EXOPOLYPHOSPHATASE.	8.51e+01
43	45	71.4	565	2	P75472	F10_ORF565 PROTEIN.	8.51e+01
44	45	71.4	980	5	Q17592	SIMILARITY TO INSULIN-	8.51e+01
45	45	71.4	1683	13	Q91741	FOURTH COMPONENT OF CO	8.51e+01

ALIGNMENTS

RESULT 1
ID O23161 PRELIMINARY; PRT; 766 AA.
AC O23161;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RECEPTOR KINASE-LIKE PROTEIN (EC 2.7.1.).
GN C7A10.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., TERRYN N., VOS P., HEIJNEN L., MEWES H.W., SCHUELLER C.,
RA CHALWATZIS N.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z99707; CAB16774.1; -.
DR MENDEL; 25486; Arath;3435;25486.
DR PFAM; PF00560; LRR; 4.
DR PFAM; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
SQ SEQUENCE 766 AA; 83775 MW; C4BD7115 CRC32;

Query Match 88.9%; Score 56; DB 10; Length 766;
Best Local Similarity 88.9%; Pred. No. 4.47e-01;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Db    337 AGIGILAVI 345
          ||||||:||
Qy      2 AGIGILTVI 10
```

RESULT 2
ID Q87090 PRELIMINARY; PRT; 478 AA.
AC Q87090;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE GLYCOPROTEIN GIII.
OS Pseudorabies virus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIANA S;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49436; BAA08414.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   478 AA;  51150 MW;  D6A143B4 CRC32;

  Query Match           82.5%;  Score 52;  DB 14;  Length 478;
  Best Local Similarity  66.7%;  Pred. No. 3.29e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    455 AGIGILAIV 463
          ||||||:::
Qy      2 AGIGILTVI 10


RESULT    3
ID   Q87089      PRELIMINARY;      PRT;   479 AA.
AC   Q87089;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YAMAGATA S-81;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49435; BAA08413.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   479 AA;  51109 MW;  A009EB9B CRC32;

  Query Match           82.5%;  Score 52;  DB 14;  Length 479;
  Best Local Similarity  66.7%;  Pred. No. 3.29e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAIV 464
          ||||||:::
Qy      2 AGIGILTVI 10


RESULT    4
ID   Q87091      PRELIMINARY;      PRT;   479 AA.
AC   Q87091;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIA3;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49437; BAA08415.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   479 AA;  51148 MW;  CC3EFF9A CRC32;

  Query Match           82.5%;  Score 52;  DB 14;  Length 479;
  Best Local Similarity  66.7%;  Pred. No. 3.29e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAIV 464
          ||||||:::
Qy      2 AGIGILTVI 10


RESULT    5
ID   O05457      PRELIMINARY;      PRT;   509 AA.
AC   O05457;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 53.2 KD PROTEIN.
GN   MTCY15F10.25.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   OLIVER K., HARRIS D.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   COLE S.T., BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; Z94121; CAB08087.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   509 AA;  53278 MW;  04302F67 CRC32;

  Query Match           79.4%;  Score 50;  DB 2;  Length 509;
  Best Local Similarity  70.0%;  Pred. No. 8.61e+00;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    172 AGGIGVLLVI 181
          |:|||:| ||
Qy      1 AAGIGILTVI 10


RESULT    6
ID   O22178      PRELIMINARY;      PRT;   773 AA.
AC   O22178;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PUTATIVE RECEPTOR PROTEIN KINASE.
GN   T20D16.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
RA   SYKES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002391; AAB87101.1; -.
```

```
DR   MENDEL; 25106; Arath;3435;25106.
DR   PFAM; PF00560; LRR; 4.
DR   PFAM; PF00069; pkinase; 1.
SQ   SEQUENCE   773 AA;  84148 MW;  83C3953B CRC32;

  Query Match           79.4%;  Score 50;  DB 10;  Length 773;
  Best Local Similarity  77.8%;  Pred. No. 8.61e+00;
  Matches     7;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    343 AGIGILALI 351
          ||||||::|
Qy      2 AGIGILTVI 10


RESULT    7
ID   Q09957      PRELIMINARY;      PRT;   295 AA.
AC   Q09957;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-FEB-1997 (TrEMBLrel. 02, Last annotation update)
DE   HYPOTHETICAL 32.3 KD PROTEIN C18A3.2 IN CHROMOSOME II.
GN   C18A3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   HALLSWORTH K.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
DR   EMBL; U28944; AAA68371.1; -.
DR   WORMPEP; C18A3.2; CE01794.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     22     42       POTENTIAL.
FT   TRANSMEM    146    166       POTENTIAL.
FT   TRANSMEM    182    202       POTENTIAL.
FT   TRANSMEM    267    287       POTENTIAL.
SQ   SEQUENCE   295 AA;  32258 MW;  D968D4E6 CRC32;

  Query Match           77.8%;  Score 49;  DB 5;  Length 295;
  Best Local Similarity  70.0%;  Pred. No. 1.38e+01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    278 AAGIGIFVII 287
          ||||||: :|
Qy      1 AAGIGILTVI 10


RESULT    8
ID   Q9X4R2      PRELIMINARY;      PRT;   344 AA.
AC   Q9X4R2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   HEAT SHOCK TRANSCRIPTION REPRESSOR HRCA.
GN   HRCA.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CP1200;
RA   KIM S.N., KIM S.W., CHOI I.H., RHEE D.K.;
RT   "hrcA in Streptococcus pneumoniae.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF117740; AAD23452.1; -.
KW   Heat shock.
SQ   SEQUENCE   344 AA;  39280 MW;  C347BD42 CRC32;

  Query Match           76.2%;  Score 48;  DB 2;  Length 344;
  Best Local Similarity  62.5%;  Pred. No. 2.20e+01;
  Matches     5;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    297 GVGILAII 304
          |:|||::|
Qy      3 GIGILTVI 10


RESULT    9
ID   O31536      PRELIMINARY;      PRT;   361 AA.
AC   O31536;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YETI PROTEIN.
GN   YETI.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99107; CAB12536.1; -.
SQ   SEQUENCE   361 AA;  41966 MW;  C3BCBED4 CRC32;

  Query Match           76.2%;  Score 48;  DB 2;  Length 361;
  Best Local Similarity  70.0%;  Pred. No. 2.20e+01;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    341 ADGIGILPLI 350
          | |||||  :|
Qy      1 AAGIGILTVI 10


RESULT   10
ID   O22848      PRELIMINARY;      PRT;   521 AA.
AC   O22848;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
```

```
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MEMBRANE TRANSPORTER D1 ISOLOG.
GN   T01O24.7.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   ROUNSLEY S.D., TSCHUDY M.M., LIN X., KETCHUM K.A., CROSBY M.L.,
RA   BRANDON R.C., SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RA   SOMERVILLE C.R., VENTER J.C.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC002335; AAB64332.1; -.
DR   MENDEL; 26517; Arath;3059;26517.
DR   PFAM; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
SQ   SEQUENCE   521 AA;  56126 MW;  B3B2F40B CRC32;

  Query Match           76.2%;  Score 48;  DB 10;  Length 521;
  Best Local Similarity  66.7%;  Pred. No. 2.20e+01;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    462 AGIAVLAVI 470
          |||::|:||
Qy      2 AGIGILTVI 10


RESULT   11
ID   P71749      PRELIMINARY;      PRT;   667 AA.
AC   P71749;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 68.3 KD PROTEIN.
GN   MTCY253.26C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   DEVLIN K., CHURCHER C.M.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; Z81368; CAB03731.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   667 AA;  68251 MW;  2C803C4F CRC32;

  Query Match           76.2%;  Score 48;  DB 2;  Length 667;
  Best Local Similarity  55.6%;  Pred. No. 2.20e+01;
  Matches     5;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    379 SGVGILVVV 387
          :|:||| |:

Qy      2 AGIGILTVI 10


RESULT   12
ID   Q89853      PRELIMINARY;      PRT;   677 AA.
AC   Q89853;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   VIRION SPIKE GLYCOPROTEIN PRECURSOR.
OS   Ebola virus (Ebo).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC   Filovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RESTON SUBTYPE, SIENA STRAIN (1992);
RX   MEDLINE; 96195018.
RA   SANCHEZ A., TRAPPIER S.G., MAHY B.W.J., PETERS C.J., NICHOL S.T.;
RT   "The virion glycoproteins of Ebola viruses are encoded in two reading
RT   frames and are expressed through transcriptional editing.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3602-3607(1996).
DR   EMBL; U23417; AAC54891.1; -.
DR   EMBL; U23416; AAC54889.1; -.
DR   PFAM; PF01611; Filo_glycop; 1.
KW   Signal.
FT   SIGNAL       1     33       POTENTIAL.
SQ   SEQUENCE   677 AA;  74523 MW;  C56FB6E0 CRC32;

  Query Match           76.2%;  Score 48;  DB 14;  Length 677;
  Best Local Similarity  70.0%;  Pred. No. 2.20e+01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    654 PAGIGIIGVI 663
          :|||||: ||
Qy      1 AAGIGILTVI 10


RESULT   13
ID   Q66799      PRELIMINARY;      PRT;   677 AA.
AC   Q66799;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   VIRION SPIKE GLYCOPROTEIN PRECURSOR.
GN   GP.
OS   Ebola virus (Ebo), and Ebola virus Reston.
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC   Filovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RESTON SUBTYPE, RESTON STRAIN;
RX   MEDLINE; 96195018.
RA   SANCHEZ A., TRAPPIER S.G., MAHY B.W.J., PETERS C.J., NICHOL S.T.;
RT   "The virion glycoproteins of Ebola viruses are encoded in two reading
RT   frames and are expressed through transcriptional editing.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3602-3607(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PENNSYLVANIA;
RX   MEDLINE; 98245155.
RA   VOLCHKOV V.E., FELDMANN H., VOLCHKOVA V.A., KLENK H.D.;
RT   "Processing of the Ebola virus glycoprotein by the proprotein
RT   convertase furin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:5762-5767(1998).
DR   EMBL; U23152; AAC54885.1; -.
DR   EMBL; AF034645; AAC24346.1; -.
DR   PFAM; PF01611; Filo_glycop; 1.
KW   Signal.
FT   SIGNAL       1     33       POTENTIAL.
SQ   SEQUENCE   677 AA;  74432 MW;  9EA5B80C CRC32;

  Query Match           76.2%;  Score 48;  DB 14;  Length 677;
  Best Local Similarity  70.0%;  Pred. No. 2.20e+01;
```

```
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    654 PAGIGIIGVI 663
          :|||||: ||
Qy      1 AAGIGILTVI 10


RESULT   14
ID   Q9ZT37      PRELIMINARY;      PRT;   808 AA.
AC   Q9ZT37;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PUTATIVE GLUTAMATE RECEPTOR.
GN   GLR1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99039497.
RA   LAM H.M., CHIU J., HSIEH M.H., MEISEL L., OLIVEIRA I.C., SHIN M.,
RA   CORUZZI G.;
RT   "Glutamate-receptor genes in plants.";
RL   Nature 396:125-126(1998).
DR   EMBL; AF079998; AAD09173.1; -.
KW   Receptor.
SQ   SEQUENCE   808 AA;  90518 MW;  C3554B89 CRC32;

  Query Match           76.2%;  Score 48;  DB 10;  Length 808;
  Best Local Similarity 100.0%;  Pred. No. 2.20e+01;
  Matches     7;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    519 GIGILTV 525
          |||||||
Qy      3 GIGILTV 9


RESULT   15
ID   Q49684      PRELIMINARY;      PRT;    91 AA.
AC   Q49684;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1996 (TrEMBLrel. 01, Last annotation update)
DE   B1496_C2_163.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ROBISON K.;
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   SMITH D.R.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   ROBISON K.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U00013; AAA17122.1; -.
SQ   SEQUENCE   91 AA;  9561 MW;  A1C4ED5D CRC32;

  Query Match           74.6%;  Score 47;  DB 2;  Length 91;
  Best Local Similarity 66.7%;  Pred. No. 3.48e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     56 AGIGVLSAI 64
          ||||:|: |
Qy      2 AGIGILTVI 10


Search completed: Fri May  5 22:20:06 2000
Job time : 89 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:17:41 2000; MasPar time 3.08 Seconds
96.985 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-18
Description: (1-10) from US09267439.pep
Perfect Score: 63
Sequence: 1 AAGIGILTVI 10

Scoring table: PAM 150
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 23.818; Variance 26.551; scale 0.897

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	63	100.0	118	1	MAR1_HUMAN	MELANOMA ANTIGEN RECOG	2.38e-03
2	52	82.5	479	1	VGLC_PRVIF	GLYCOPROTEIN GIII PREC	1.03e+00
3	50	79.4	461	1	YXCC_BACSU	HYPOTHETICAL METABOLIT	2.87e+00
4	49	77.8	345	1	HRCA_STRMU	HEAT-INDUCIBLE TRANSCR	4.74e+00
5	49	77.8	885	1	YDGH_BACSU	PUTATIVE MEMBRANE PROT	4.74e+00
6	48	76.2	231	1	YSA1_YEAST	YSA1 PROTEIN.	7.76e+00
7	48	76.2	313	1	Y4TP_RHISN	PROBABLE PEPTIDE ABC T	7.76e+00
8	48	76.2	487	1	Y346_MYCTU	HYPOTHETICAL 52.2 KD T	7.76e+00
9	48	76.2	633	1	Y561_HAEIN	HYPOTHETICAL PROTEIN H	7.76e+00
10	48	76.2	1331	1	CYAB_LEIDO	RECEPTOR-TYPE ADENYLAT	7.76e+00
11	47	74.6	101	1	ATPL_SULAC	MEMBRANE-ASSOCIATED AT	1.26e+01
12	47	74.6	308	1	MENA_HAEIN	1,4-DIHYDROXY-2-NAPHTH	1.26e+01
13	47	74.6	503	1	SECD_HELPY	PROTEIN-EXPORT MEMBRAN	1.26e+01
14	47	74.6	526	1	SECD_HELPJ	PROTEIN-EXPORT MEMBRAN	1.26e+01
15	46	73.0	201	1	Y760_PYRHO	HYPOTHETICAL PROTEIN P	2.03e+01
16	46	73.0	337	1	OPSX_HUMAN	VISUAL PIGMENT-LIKE RE	2.03e+01
17	46	73.0	440	1	UGTC_CAEEL	PUTATIVE UDP-GLUCURONO	2.03e+01
18	46	73.0	977	1	YDG8_SCHPO	HYPOTHETICAL 111.4 KD	2.03e+01
19	46	73.0	1109	1	CYGD_CANFA	RETINAL GUANYLYL CYCLA	2.03e+01
20	45	71.4	235	1	YQGA_ECOLI	HYPOTHETICAL 24.6 KD P	3.24e+01
21	45	71.4	291	1	Y4TQ_RHISN	PROBABLE PEPTIDE ABC T	3.24e+01
22	45	71.4	385	1	AMPC_LYSLA	BETA-LACTAMASE PRECURS	3.24e+01
23	45	71.4	566	1	NARQ_ECOLI	NITRATE/NITRITE SENSOR	3.24e+01
24	44	69.8	278	1	YQHA_BACSU	HYPOTHETICAL 31.8 KD P	5.13e+01
25	44	69.8	289	1	HBD_CLOTS	3-HYDROXYBUTYRYL-COA D	5.13e+01
26	44	69.8	394	1	FTSZ_AZOVI	CELL DIVISION PROTEIN	5.13e+01
27	44	69.8	456	1	GLMU_ECOLI	UDP-N-ACETYLGLUCOSAMIN	5.13e+01
28	44	69.8	610	1	FIMB_DICDI	FIMBRIN.	5.13e+01
29	44	69.8	611	1	YD3M_HERAU	HYPOTHETICAL 68.4 KD P	5.13e+01
30	44	69.8	635	1	XYND_PAEPO	ENDO-1,4-BETA-XYLANASE	5.13e+01
31	44	69.8	1091	1	NCA1_CHICK	NEURAL CELL ADHESION M	5.13e+01
32	43	68.3	103	1	SYB3_MOUSE	SYNAPTOBREVIN 3 (CELLU	8.05e+01
33	43	68.3	132	1	ATPE_ARATH	ATP SYNTHASE EPSILON C	8.05e+01
34	43	68.3	216	1	FLA2_METVO	FLAGELLIN B2 PRECURSOR	8.05e+01
35	43	68.3	218	1	FLA1_METVO	FLAGELLIN B1 PRECURSOR	8.05e+01
36	43	68.3	220	1	YO69_CAEEL	HYPOTHETICAL 23.9 KD P	8.05e+01
37	43	68.3	222	1	FLA2_METVA	FLAGELLIN B2 PRECURSOR	8.05e+01
38	43	68.3	342	1	AQPL_HUMAN	AQUAPORIN-7 LIKE (AQUA	8.05e+01
39	43	68.3	366	1	YI21_SYNY3	HYPOTHETICAL 39.0 KD P	8.05e+01
40	43	68.3	394	1	FTSW_HAEIN	CELL DIVISION PROTEIN	8.05e+01
41	43	68.3	404	1	SGAA_HYPME	SERINE--GLYOXYLATE AMI	8.05e+01
42	43	68.3	493	1	ACHE_MOUSE	ACETYLCHOLINE RECEPTOR	8.05e+01
43	43	68.3	536	1	FLIF_CAUCR	FLAGELLAR M-RING PROTE	8.05e+01
44	43	68.3	590	1	MDLA_ECOLI	MULTIDRUG RESISTANCE-L	8.05e+01
45	43	68.3	1108	1	CYGE_RAT	GUANYLYL CYCLASE GC-E	8.05e+01

ALIGNMENTS

```
RESULT    1
ID   MAR1_HUMAN     STANDARD;      PRT;   118 AA.
AC   Q16655;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MELANOMA ANTIGEN RECOGNIZED BY T-CELLS 1 (MART-1) (MELAN-A PROTEIN)
DE   (ANTIGEN SK29-AA) (ANTIGEN LB39-AA).
GN   MLANA OR MART1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RX   MEDLINE; 94224770.
RA   KAWAKAMI Y., ELIYAHU S., DELGADO C.H., ROBBINS P.F., RIVOLTINI L.,
RA   TOPALIAN S.L., MIKI T., ROSENBERG S.A.;
RT   "Cloning of the gene coding for a shared human melanoma antigen
RT   recognized by autologous T cells infiltrating into tumor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3515-3519(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94275389.
RA   COULIE P.G., BRICHARD V., VAN PEL A., WOELFEL T., SCHNEIDER J.,
RA   TRAVERSARI C., MATTEI S., DE PLAEN E., LURQUIN C., SZIKORA J.-P.,
RA   RENAULD J.-C., BOON T.;
RT   "A new gene coding for a differentiation antigen recognized by
RT   autologous cytolytic T lymphocytes on HLA-A2 melanomas.";
RL   J. Exp. Med. 180:35-42(1994).
CC   -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO MELANOMA AND
CC       MELANOCYTE CELL LINES AND RETINA.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U06452; AAA19238.1; -.
DR   EMBL; U06654; AAA20389.1; -.
KW   Antigen; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
SQ   SEQUENCE   118 AA;  13157 MW;  DFE2CFC6 CRC32;
```

```
  Query Match          100.0%;  Score 63;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 2.38e-03;
  Matches   10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 AAGIGILTVI 36
          ||||||||||
Qy      1 AAGIGILTVI 10


RESULT    2
ID   VGLC_PRVIF     STANDARD;      PRT;   479 AA.
AC   P06024;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   GLYCOPROTEIN GIII PRECURSOR.
OS   Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86200375.
RA   ROBBINS A.K., WATSON R.J., WHEALY M.E., HAYS W.W., ENQUIST L.W.;
RT   "Characterization of a pseudorabies virus glycoprotein gene with
RT   homology to herpes simplex virus type 1 and type 2 glycoprotein C.";
RL   J. Virol. 58:339-347(1986).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN C.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN GENE SUPERFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M12778; AAA47464.1; -.
DR   PIR; A26097; VGBEPB.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    479       GLYCOPROTEIN GIII.
FT   CARBOHYD     40     40       POTENTIAL.
FT   CARBOHYD     84     84       POTENTIAL.
FT   CARBOHYD    169    169       POTENTIAL.
FT   CARBOHYD    192    192       POTENTIAL.
FT   CARBOHYD    220    220       POTENTIAL.
FT   CARBOHYD    228    228       POTENTIAL.
FT   CARBOHYD    285    285       POTENTIAL.
FT   CARBOHYD    302    302       POTENTIAL.
SQ   SEQUENCE   479 AA;  51206 MW;  42EE5703 CRC32;

  Query Match          82.5%;  Score 52;  DB 1;  Length 479;
  Best Local Similarity 66.7%;  Pred. No. 1.03e+00;
  Matches    6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAIV 464
          ||||||:::
Qy      2 AGIGILTVI 10


RESULT    3
ID   YXCC_BACSU     STANDARD;      PRT;   461 AA.
AC   P46333; O32289;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN IOLS-HTPG INTERGENIC
DE   REGION.
GN   YXCC OR SS92BR.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE; 96093926.
RA   YOSHIDA K.-I., SEKI S., FUJIMURA M., MIWA Y., FUJITA Y.;
RT   "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RT   genome between the gnt and iol operons.";
RL   DNA Res. 2:61-69(1995).
RN   [2]
RP   REVISIONS.
RA   FUJITA Y., SHIBAYAMA T., ISHIO I., AOYAMA D., YOSHIDA K.-I.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB005554; BAA21604.1; -.
DR   EMBL; Z99124; CAB16017.1; -.
DR   SUBTILIST; BG11360; YXCC.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR   PFAM; PF00083; sugar_tr; 1.
KW   Hypothetical protein; Transport; Transmembrane.
FT   TRANSMEM     15     35       POTENTIAL.
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM     77     97       POTENTIAL.
FT   TRANSMEM    105    125       POTENTIAL.
FT   TRANSMEM    140    160       POTENTIAL.
FT   TRANSMEM    164    184       POTENTIAL.
FT   TRANSMEM    242    262       POTENTIAL.
FT   TRANSMEM    281    301       POTENTIAL.
FT   TRANSMEM    309    329       POTENTIAL.
FT   TRANSMEM    342    362       POTENTIAL.
FT   CONFLICT    400    401       RP -> SA (IN REF. 3).
SQ   SEQUENCE   461 AA;  50235 MW;  D47572C9 CRC32;

  Query Match          79.4%;  Score 50;  DB 1;  Length 461;
  Best Local Similarity 87.5%;  Pred. No. 2.87e+00;
  Matches    7;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    285 GIGILNVI 292
          ||||| ||
Qy      3 GIGILTVI 10


RESULT    4
ID   HRCA_STRMU     STANDARD;      PRT;   345 AA.
AC   O06940;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HEAT-INDUCIBLE TRANSCRIPTION REPRESSOR HRCA.
GN   HRCA.
OS   Streptococcus mutans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Streptococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GS-5;
RX   MEDLINE; 97426041.
RA   JAYARAMAN G.C., PENDERS J.E., BURNE R.A.;
```

```
RT   "Transcriptional analysis of the Streptococcus mutans hrcA, grpE and
RT   dnaK genes and regulation of expression in response to heat shock and
RT   environmental acidification.";
RL   Mol. Microbiol. 25:329-341(1997).
CC   -!- FUNCTION: NEGATIVE REGULATOR OF CLASS I HEAT SHOCK GENES (GRPE-
CC       DNAK-DNAJ AND GROELS OPERONS). PREVENTS HEAT-SHOCK INDUCTION OF
CC       THESE OPERONS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HRCA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U78296; AAC45610.1; -.
DR   PFAM; PF01628; HrcA; 1.
KW   Transcription regulation; Repressor; Heat shock.
SQ   SEQUENCE   345 AA;  39306 MW;  A0B065C2 CRC32;

  Query Match           77.8%;  Score 49;  DB 1;  Length 345;
  Best Local Similarity  87.5%;  Pred. No. 4.74e+00;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    298 GFGILTVI 305
          |:||||||
Qy      3 GIGILTVI 10


RESULT    5
ID   YDGH_BACSU     STANDARD;      PRT;   885 AA.
AC   P96706;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN YDGH.
GN   YDGH.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KASAHARA Y., NAKAI S., LEE S., SADAIE Y., OGASAWARA N.;
RT   "A 148 kbp sequence of the region between 35 and 47 degree of the
RT   Bacillus subtilis genome.";
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB001488; BAA19398.1; -.
DR   EMBL; Z99106; CAB12372.1; -.
DR   EMBL; Z99107; CAB12384.1; -.
DR   SUBTILIST; BG12175; YDGH.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
FT   TRANSMEM    202    222       POTENTIAL.
FT   TRANSMEM    227    247       POTENTIAL.
FT   TRANSMEM    278    298       POTENTIAL.
FT   TRANSMEM    304    324       POTENTIAL.
FT   TRANSMEM    354    374       POTENTIAL.
FT   TRANSMEM    716    736       POTENTIAL.
FT   TRANSMEM    740    760       POTENTIAL.
FT   TRANSMEM    772    792       POTENTIAL.
FT   TRANSMEM    817    837       POTENTIAL.
FT   TRANSMEM    847    867       POTENTIAL.
SQ   SEQUENCE   885 AA;  95488 MW;  4106171D CRC32;

  Query Match           77.8%;  Score 49;  DB 1;  Length 885;
  Best Local Similarity  60.0%;  Pred. No. 4.74e+00;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    305 AVGVGILMII 314
          | |:||| :|
Qy      1 AAGIGILTVI 10


RESULT    6
ID   YSA1_YEAST     STANDARD;      PRT;   231 AA.
AC   Q01976;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   YSA1 PROTEIN.
GN   YSA1 OR YBR111C OR YBR0907.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 95208357.
RA   MANNHAUPT G., STUCKA R., EHNLE S., VETTER I., FELDMANN H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
RN   [2]
RP   SEQUENCE OF 1-47 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 92327848.
RA   MANNHAUPT G., STUCKA R., EHNLE S., VETTER I., FELDMANN H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
CC   -!- SIMILARITY: STRONG, TO B.SUBTILIS YQKG.
CC   -!- SIMILARITY: TO PROTEINS WITH A CORE MUTT DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35980; CAA85068.1; -.
DR   EMBL; X78993; CAA55614.1; -.
DR   EMBL; X66247; CAA46972.1; -.
DR   PIR; S44691; S44691.
DR   SGD; L0002551; YSA1.
DR   PROSITE; PS00893; MUTT; 1.
DR   PFAM; PF00293; mutT; 1.
FT   DOMAIN      112    145       MUTT-LIKE.
SQ   SEQUENCE   231 AA;  26087 MW;  49A2D6CB CRC32;

  Query Match           76.2%;  Score 48;  DB 1;  Length 231;
  Best Local Similarity  75.0%;  Pred. No. 7.76e+00;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     79 GIGILTIL 86
          ||||||::
Qy      3 GIGILTVI 10


RESULT    7
```

```
ID   Y4TP_RHISN     STANDARD;      PRT;   313 AA.
AC   Q53191;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROBABLE PEPTIDE ABC TRANSPORTER PERMEASE PROTEIN Y4TP.
GN   Y4TP.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97305956.
RA   FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA   PERRET X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
RN   [2]
RP   SEQUENCE OF 107-313 FROM N.A.
RX   MEDLINE; 96389014.
RA   FREIBERG C., PERRET X., BROUGHTON W.J., ROSENTHAL A.;
RT   "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.
RT   NGR234 using dye terminators and a thermostable 'sequenase': a
RT   beginning.";
RL   Genome Res. 6:590-600(1996).
CC   -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM Y4TOPQRS FOR A PEPTIDE. PROBABLY RESPONSIBLE FOR THE
CC       TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPPBC
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000098; AAB91869.1; -.
DR   EMBL; Z68203; CAA92398.1; -.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
DR   PFAM; PF00528; BPD_transp; 1.
KW   Hypothetical protein; Transport; Amino-acid transport; Transmembrane;
KW   Inner membrane; Plasmid.
FT   TRANSMEM      9     29       POTENTIAL.
FT   TRANSMEM    101    121       POTENTIAL.
FT   TRANSMEM    137    157       POTENTIAL.
FT   TRANSMEM    177    197       POTENTIAL.
FT   TRANSMEM    236    256       POTENTIAL.
FT   TRANSMEM    280    300       POTENTIAL.
SQ   SEQUENCE   313 AA;  34042 MW;  31F3F704 CRC32;

  Query Match           76.2%;  Score 48;  DB 1;  Length 313;
  Best Local Similarity  77.8%;  Pred. No. 7.76e+00;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    235 AGVPILTVI 243
          ||: |||||
Qy      2 AGIGILTVI 10


RESULT    8
ID   Y346_MYCTU     STANDARD;      PRT;   487 AA.
AC   O06297;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 52.2 KD TRANSPORT PROTEIN RV0346C.
GN   RV0346C OR MTCY13E10.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 98295987.
RA   COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA   GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA   BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA   DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA   HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA   OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA   RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQARES R., SULSTON J.E.,
RA   TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95324; CAB08578.1; -.
DR   PROSITE; PS00218; AMINO_ACID_PERMEASE; 1.
DR   PFAM; PF00324; aa_permeases; 1.
KW   Hypothetical protein; Transport; Amino-acid transport; Transmembrane.
FT   TRANSMEM     26     46       POTENTIAL.
FT   TRANSMEM     50     70       POTENTIAL.
FT   TRANSMEM     98    118       POTENTIAL.
FT   TRANSMEM    133    153       POTENTIAL.
FT   TRANSMEM    163    183       POTENTIAL.
FT   TRANSMEM    214    234       POTENTIAL.
FT   TRANSMEM    256    276       POTENTIAL.
FT   TRANSMEM    290    310       POTENTIAL.
FT   TRANSMEM    341    361       POTENTIAL.
FT   TRANSMEM    369    389       POTENTIAL.
FT   TRANSMEM    414    434       POTENTIAL.
FT   TRANSMEM    440    460       POTENTIAL.
SQ   SEQUENCE   487 AA;  52194 MW;  64BBBCCD CRC32;

  Query Match           76.2%;  Score 48;  DB 1;  Length 487;
  Best Local Similarity  60.0%;  Pred. No. 7.76e+00;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    351 TAGIGLLGII 360
          :||||:| :|
Qy      1 AAGIGILTVI 10


RESULT    9
ID   Y561_HAEIN     STANDARD;      PRT;   633 AA.
AC   P44016;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL PROTEIN HI0561/560.
GN   HI0561/560.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
```

```
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   REVISIONS.
RA   WHITE O., CLAYTON R.A., KERLAVAGE A.R., FLEISCHMANN R.D.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32737; AAC22215.1; -.
DR   TIGR; HI0561; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     45     65       POTENTIAL.
FT   TRANSMEM     70     90       POTENTIAL.
FT   TRANSMEM    128    148       POTENTIAL.
FT   TRANSMEM    180    200       POTENTIAL.
FT   TRANSMEM    230    250       POTENTIAL.
FT   TRANSMEM    281    301       POTENTIAL.
FT   TRANSMEM    311    331       POTENTIAL.
FT   TRANSMEM    345    365       POTENTIAL.
FT   TRANSMEM    379    399       POTENTIAL.
FT   TRANSMEM    420    440       POTENTIAL.
FT   TRANSMEM    483    503       POTENTIAL.
FT   TRANSMEM    515    535       POTENTIAL.
FT   TRANSMEM    564    584       POTENTIAL.
FT   TRANSMEM    604    624       POTENTIAL.
SQ   SEQUENCE   633 AA;  66616 MW;  A9409DDD CRC32;

  Query Match           76.2%;  Score 48;  DB 1;  Length 633;
  Best Local Similarity  66.7%;  Pred. No. 7.76e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    347 SGIGIISVI 355
          :||||::||
Qy      2 AGIGILTVI 10


RESULT   10
ID   CYAB_LEIDO     STANDARD;      PRT;  1331 AA.
AC   Q25263;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RECEPTOR-TYPE ADENYLATE CYCLASE B (EC 4.6.1.1) (ATP PYROPHOSPHATE-
DE   LYASE) (ADENYLYL CYCLASE).
GN   RAC-B.
OS   Leishmania donovani.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1S SUDANESE;
RX   MEDLINE; 95340554.
RA   SANCHEZ M.A., ZEOLI D., KLAMO E.M., KAVANAUGH M.P., LANDFEAR S.M.;
RT   "A family of putative receptor-adenylate cyclases from Leishmania
RT   donovani.";
RL   J. Biol. Chem. 270:17551-17558(1995).
CC   -!- FUNCTION: COULD ACT AS A RECEPTOR FOR A UNKNOWN LIGAND.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE INSECT STAGE (PROMASTIGOTE)
CC       BUT NOT IN THE MAMMALIAN HOST STAGE OF THE PARASITE LIFE CYCLE.
CC   -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U17043; AAA74999.1; -.
DR   PFAM; PF00211; guanylate_cyc; 1.
KW   Lyase; cAMP synthesis; Transmembrane; Receptor; Glycoprotein.
FT   DOMAIN        1     33       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     34     54       POTENTIAL.
FT   DOMAIN       55    898       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    899    919       POTENTIAL.
FT   DOMAIN      920   1331       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    429    429       POTENTIAL.
FT   CARBOHYD    558    558       POTENTIAL.
FT   CARBOHYD    574    574       POTENTIAL.
FT   CARBOHYD    657    657       POTENTIAL.
SQ   SEQUENCE  1331 AA;  144162 MW;  CCC01FC9 CRC32;

  Query Match           76.2%;  Score 48;  DB 1;  Length 1331;
  Best Local Similarity  77.8%;  Pred. No. 7.76e+00;
  Matches     7;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    902 AGIALLTVI 910
          |||::||||
Qy      2 AGIGILTVI 10


RESULT   11
ID   ATPL_SULAC     STANDARD;      PRT;  101 AA.
AC   P23040;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MEMBRANE-ASSOCIATED ATPASE C CHAIN (EC 3.6.1.34) (SUL-ATPASE
DE   PROTEOLIPID CHAIN).
GN   ATPP.
OS   Sulfolobus acidocaldarius.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 89214142.
RA   DENDA K., KONISHI J., OSHIMA T., DATE T., YOSHIDA M.;
RT   "A gene encoding the proteolipid subunit of Sulfolobus acidocaldarius
RT   ATPase complex.";
RL   J. Biol. Chem. 264:7119-7121(1989).
CC   -!- FUNCTION: THE C CHAIN IS A PROTEOLIPID, AND ONE OF THE MEMBRANOUS
CC       SUBUNITS OF THE THE NONENZYMATIC COMPONENT OF THE SUL-ATPASE
CC       COMPLEX.
CC   -!- SUBUNIT: SUL-ATPASE IS COMPOSED OF SIX (OR FIVE ?) SUBUNITS:
CC       ALPHA, BETA, DELTA, GAMMA, C (PROTEOLIPID), AND POSSIBLY EPSILON.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE V-ATPASE PROTEOLIPID SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ---------------------------------------------------------------------
DR   EMBL; J04740; AAA72703.1; -.
DR   PIR; A33351; A33351.
DR   HSSP; P00138; 1CGN.
DR   PFAM; PF00137; ATP-synt_C; 1.
KW   Hydrogen ion transport; Lipid-binding; Transmembrane.
FT   TRANSMEM      5      25       POTENTIAL.
FT   TRANSMEM     37      57       POTENTIAL.
FT   TRANSMEM     75      95       POTENTIAL.
SQ   SEQUENCE   101 AA;  10362 MW;  1DC8C74D CRC32;

  Query Match           74.6%;  Score 47;  DB 1;  Length 101;
  Best Local Similarity  87.5%;  Pred. No. 1.26e+01;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 AAGIGVLT 66
          |||||:||
Qy      1 AAGIGILT 8


RESULT   12
ID   MENA_HAEIN     STANDARD;      PRT;   308 AA.
AC   P44739;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   1,4-DIHYDROXY-2-NAPHTHOATE OCTAPRENYLTRANSFERASE (EC 2.5.1.-) (DHNA-
DE   OCTAPRENYLTRANSFERASE).
GN   MENA OR HI0509.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20;
RX   MEDLINE; 95350630.
RA   FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA   KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA   MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA   SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA   WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA   UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA   FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA   GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA   VENTER J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: CONVERSION OF 1,4-DIHYDROXY-2-NAPHTHOATE (DHNA) TO
CC       DIMETHYLMENAQUINONE (DMK). ATTACHES OCTAPRENYLPYROPHOSPHATE, A
CC       MEMBRANE-BOUND 40-CARBON SIDE CHAIN TO DHNA. THE CONVERSION OF
CC       DHNA TO DMK PROCEEDS IN THREE STAGES: THE REMOVAL OF THE CARBOXYL
CC       GROUP OF DHNA AS CO2, THE ATTACHMENT OF THE ISOPRENOID SIDE CHAIN,
CC       AND A QUINOL-TO-QUINONE OXIDATION, WHICH IS THOUGHT TO BE
CC       SPONTANEOUS (BY SIMILARITY).
CC   -!- PATHWAY: MENAQUINONE BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MENA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32732; AAC22167.1; -.
DR   TIGR; HI0509; -.
KW   Menaquinone biosynthesis; Transferase; Transmembrane; Inner membrane.
FT   TRANSMEM      22      42       POTENTIAL.
FT   TRANSMEM      47      67       POTENTIAL.
FT   TRANSMEM     101     121       POTENTIAL.
FT   TRANSMEM     129     149       POTENTIAL.
FT   TRANSMEM     153     173       POTENTIAL.
FT   TRANSMEM     186     206       POTENTIAL.
FT   TRANSMEM     235     255       POTENTIAL.
FT   TRANSMEM     286     306       POTENTIAL.
SQ   SEQUENCE   308 AA;  33345 MW;  090B2655 CRC32;

  Query Match           74.6%;  Score 47;  DB 1;  Length 308;
  Best Local Similarity  55.6%;  Pred. No. 1.26e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    132 AGLGILAIV 140
          ||:|||:::
Qy      2 AGIGILTVI 10


RESULT   13
ID   SECD_HELPY     STANDARD;      PRT;   503 AA.
AC   O26074;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   PROTEIN-EXPORT MEMBRANE PROTEIN SECD.
GN   SECD OR HP1550.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE; 97394467.
RA   TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA   FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA   NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA   LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA   MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA   BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA   COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA   HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA   VENTER J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC   -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000652; AAD08588.1; -.
DR   TIGR; HP1550; -.
KW   Protein transport; Translocation; Transmembrane; Membrane.
FT   TRANSMEM     199     219       POTENTIAL.
FT   TRANSMEM     334     354       POTENTIAL.
FT   TRANSMEM     357     377       POTENTIAL.
FT   TRANSMEM     383     403       POTENTIAL.
FT   TRANSMEM     456     476       POTENTIAL.
SQ   SEQUENCE   503 AA;  54247 MW;  9A76592C CRC32;

  Query Match           74.6%;  Score 47;  DB 1;  Length 503;
  Best Local Similarity  60.0%;  Pred. No. 1.26e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Db    460 TTGIGILASI 469
          ::||||||: |
Qy      1 AAGIGILTVI 10


RESULT   14
ID   SECD_HELPJ     STANDARD;      PRT;   526 AA.
AC   Q9ZJ66;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   PROTEIN-EXPORT MEMBRANE PROTEIN.
GN   SECD.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99120557.
RA   ALM R.A., LING L.-S.L., MOIR D.T., KING B.L., BROWN E.D., DOIG P.C.,
RA   SMITH D.R., NOONAN B., GUILD B.C., DEJONGE B.L., CARMEL G.,
RA   TUMMINO P.J., CARUSO A., URIA-NICKELSEN M., MILLS D.M., IVES C.,
RA   GIBSON R., MERBERG D., MILLS S.D., JIANG Q., TAYLOR D.E., VOVIS G.F.,
RA   TRUST T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC   -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS
CC       WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001567; AAD07024.1; -.
KW   Protein transport; Translocation; Transmembrane; Membrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM    356    376       POTENTIAL.
FT   TRANSMEM    379    399       POTENTIAL.
FT   TRANSMEM    453    473       POTENTIAL.
FT   TRANSMEM    478    498       POTENTIAL.
SQ   SEQUENCE   526 AA;  56796 MW;  FCDF8FA9 CRC32;

  Query Match           74.6%;  Score 47;  DB 1;  Length 526;
  Best Local Similarity  60.0%;  Pred. No. 1.26e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    482 TTGIGILASI 491
          ::||||||: |
Qy      1 AAGIGILTVI 10


RESULT   15
ID   Y760_PYRHO     STANDARD;      PRT;   201 AA.
AC   O58499;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN PH0760.
GN   PH0760 OR PHCI026.
OS   Pyrococcus horikoshii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OT3;
RX   MEDLINE; 98344137.
RA   KAWARABAYASI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA   YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOYAMA A., NAGAI Y.,
RA   SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,
RA   FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA   AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA   KIKUCHI H.;
RT   "Complete sequence and gene organization of the genome of a hyper-
RT   thermophilic archaebacterium, Pyrococcus horikoshii OT3.";
RL   DNA Res. 5:55-76(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE UPF0056 (MARC) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP000003; BAA29851.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     49     69       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM    111    131       POTENTIAL.
FT   TRANSMEM    140    160       POTENTIAL.
FT   TRANSMEM    181    201       POTENTIAL.
SQ   SEQUENCE   201 AA;  21592 MW;  97675186 CRC32;

  Query Match           73.0%;  Score 46;  DB 1;  Length 201;
  Best Local Similarity  60.0%;  Pred. No. 2.03e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    138 SPGIVILTII 147
          ::|| |||:|
Qy      1 AAGIGILTVI 10


Search completed: Fri May  5 22:18:20 2000
Job time : 39 secs.
```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:16:24 2000; MasPar time 4.76 Seconds
99.131 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-18
Description: (1-10) from US09267439.pep
Perfect Score: 63
Sequence: 1 AAGIGILTVI 10

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 23.344; Variance 29.628; scale 0.788

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	63	100.0	118	2	A55253	melanoma antigen MART	1.70e-02
2	52	82.5	479	1	VGBEPB	glycoprotein gIII pre	3.90e+00
3	50	79.4	461	2	D70073	metabolite transport	9.80e+00
4	50	79.4	509	2	H70597	probable membrane pro	9.80e+00
5	50	79.4	773	2	T00502	protein kinase homolo	9.80e+00
6	49	77.8	295	2	T15544	hypothetical protein	1.54e+01
7	49	77.8	885	2	B69783	transporter homolog y	1.54e+01
8	48	76.2	231	1	S48276	YSA1 protein - yeast	2.40e+01
9	48	76.2	361	2	F69798	conserved hypothetica	2.40e+01
10	48	76.2	487	2	C70574	probable aroP2 protei	2.40e+01
11	48	76.2	667	2	F70682	probable membrane pro	2.40e+01
12	47	74.6	91	2	S72755	B1496_C2_163 protein	3.71e+01
13	47	74.6	101	2	A33351	H+-transporting ATP s	3.71e+01
14	47	74.6	165	2	C70959	hypothetical protein	3.71e+01
15	47	74.6	250	2	A69843	hypothetical protein	3.71e+01
16	47	74.6	308	2	H64153	hypothetical protein	3.71e+01
17	47	74.6	503	1	F64713	protein-export membra	3.71e+01
18	47	74.6	526	2	D71805	protein-export membra	3.71e+01
19	46	73.0	201	2	A71124	hypothetical protein	5.71e+01
20	46	73.0	271	2	B72270	conserved hypothetica	5.71e+01
21	46	73.0	287	2	F72780	hypothetical protein	5.71e+01
22	46	73.0	310	2	D75202	dipeptide abc transpo	5.71e+01
23	46	73.0	637	2	S35221	globulin Beg1 precurs	5.71e+01
24	46	73.0	779	2	G70027	conserved hypothetica	5.71e+01
25	45	71.4	98	2	D72106	hypothetical protein	8.72e+01
26	45	71.4	217	2	G70378	hypothetical protein	8.72e+01
27	45	71.4	235	2	E65082	hypothetical protein	8.72e+01
28	45	71.4	339	2	S62369	methylcobalamin--coen	8.72e+01
29	45	71.4	385	2	S54103	beta-lactamase (EC 3.	8.72e+01
30	45	71.4	420	2	S59131	Kan-1 protein - rat	8.72e+01
31	45	71.4	565	2	S73854	hypothetical protein	8.72e+01
32	45	71.4	566	2	D65022	nitrate/nitrite senso	8.72e+01
33	45	71.4	585	2	A69829	ABC transporter (ATP-	8.72e+01
34	44	69.8	107	1	C69166	conserved hypothetica	1.32e+02
35	44	69.8	110	2	G72252	hypothetical protein	1.32e+02
36	44	69.8	264	2	A71367	probable holocytochro	1.32e+02
37	44	69.8	278	1	D69958	conserved hypothetica	1.32e+02
38	44	69.8	394	2	A55045	probable 3-hydroxyacy	1.32e+02
39	44	69.8	398	1	G71118	hypothetical protein	1.32e+02
40	44	69.8	456	2	C65176	glmU protein - Escher	1.32e+02
41	44	69.8	486	2	F69762	transporter homolog y	1.32e+02
42	44	69.8	556	2	S76624	integral membrane pro	1.32e+02
43	44	69.8	620	2	H69382	ABC transporter, ATP-	1.32e+02
44	44	69.8	635	2	S19011	endo-1,4-beta-xylanas	1.32e+02
45	44	69.8	1091	1	IJCHNL	neural cell adhesion	1.32e+02

ALIGNMENTS

RESULT 1
ENTRY A55253 #type complete
TITLE melanoma antigen MART-1 - human
ALTERNATE_NAMES melan-A protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 10-Sep-1997
ACCESSIONS A55253; I38506
REFERENCE A55253
#authors Kawakami, Y.; Eliyahu, S.; Delgado, C.H.; Robbins, P.F.; Rivoltini, L.; Topalian, S.L.; Miki, T.; Rosenberg, S.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:3515-3519
#title Cloning of the gene coding for a shared human melanoma antigen recognized by autologous T cells infiltrating into tumor.
#cross-references MUID:94224770
#accession A55253
##status preliminary
##molecule_type mRNA
##residues 1-118 ##label KAW
##cross-references GB:U06452; NID:g476131; PID:g476132
REFERENCE I38506
#authors Coulie, P.G.; Brichard, V.; Van Pel, A.; Wolfel, T.; Schneider, J.; Traversari, C.; Mattei, S.; De Plaen, E.; Lurquin, C.; Szikora, J.P.; Renauld, J.; Boon, T.
#journal J. Exp. Med. (1994) 180:35-42
#title A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas [see comments].
#cross-references MUID:94275389
#accession I38506
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-118 ##label RES
##cross-references EMBL:U06654; NID:g517022; PID:g517023
GENETICS
#gene GDB:MLANA
##cross-references GDB:358979
#map_position 17q21-17q24
SUMMARY #length 118 #molecular-weight 13157 #checksum 3535

Query Match 100.0%; Score 63; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 1.70e-02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 27 AAGIGILTVI 36
||||||||||

```
Qy      1 AAGIGILTVI 10


RESULT    2
ENTRY             VGBEPB     #type complete
TITLE             glycoprotein gIII precursor - suid herpesvirus 1
ORGANISM          #formal_name suid herpesvirus 1
DATE              30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
                    16-Jul-1999
ACCESSIONS        A26097
REFERENCE         A26097
   #authors       Robbins, A.K.; Watson, R.J.; Whealy, M.E.; Hays, W.W.;
                    Enquist, L.W.
   #journal       J. Virol. (1986) 58:339-347
   #title         Characterization of a Pseudorabies virus glycoprotein gene
                    with homology to herpes simplex virus type 1 and type 2
                    glycoprotein C.
   #cross-references MUID:86200375
   #accession     A26097
      ##molecule_type DNA
      ##residues      1-479 ##label ROB
      ##cross-references GB:M12778; NID:g334049; PIDN:AAA47464.1; PID:g334050
      ##experimental_source strain Becker
CLASSIFICATION    #superfamily herpesvirus glycoprotein F
KEYWORDS          glycoprotein
FEATURE
   1-22                #domain signal sequence #status predicted #label SIG\
   23-479              #product glycoprotein gIII #status predicted #label GPG\
   40,84,169,192,220,
   228,285,302         #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY           #length 479  #molecular-weight 51206  #checksum 1630

  Query Match           82.5%;  Score 52;  DB 1;  Length 479;
  Best Local Similarity  66.7%;  Pred. No. 3.90e+00;
  Matches     6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAIV 464
          ||||||:::
Qy      2 AGIGILTVI 10


RESULT    3
ENTRY             D70073     #type complete
TITLE             metabolite transport protein homolog yxcC - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    24-Sep-1999
ACCESSIONS        D70073
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #cross-references MUID:98044033
   #accession     D70073
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues      1-461 ##label KUN
      ##cross-references GB:Z99124; GB:AL009126; NID:g2636442;
                           PIDN:CAB16017.1; PID:e1184706; PID:g2636527
      ##experimental_source strain 168
GENETICS
   #gene          yxcC
CLASSIFICATION    #superfamily glucose transport protein
SUMMARY           #length 461  #molecular-weight 50140  #checksum 8642

  Query Match           79.4%;  Score 50;  DB 2;  Length 461;
  Best Local Similarity  87.5%;  Pred. No. 9.80e+00;
  Matches     7;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    285 GIGILNVI 292
          ||||| ||
Qy      3 GIGILTVI 10


RESULT    4
ENTRY             H70597     #type complete
TITLE             probable membrane protein - Mycobacterium tuberculosis
                    (strain H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    17-Jul-1998
ACCESSIONS        H70597
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537-544
   #title         Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
   #cross-references MUID:98295987
   #accession     H70597
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues      1-509 ##label COL
      ##cross-references GB:Z94121; GB:AL123456; NID:g3261736; PID:e312290;
                           PID:g1944601
      ##experimental_source strain H37Rv
GENETICS
   #gene          Rv3887c
SUMMARY           #length 509  #molecular-weight 53278  #checksum 6762
```

```
  Query Match           79.4%;  Score 50;  DB 2;  Length 509;
  Best Local Similarity  70.0%;  Pred. No. 9.80e+00;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    172 AGGIGVLLVI 181
          |:|||:| ||
Qy      1 AAGIGILTVI 10


RESULT    5
ENTRY             T00502     #type complete
TITLE             protein kinase homolog T20D16.7 - Arabidopsis thaliana
ORGANISM          #formal_name Arabidopsis thaliana #common_name mouse-ear
                    cress
DATE              01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
                    23-Apr-1999
ACCESSIONS        T00502
REFERENCE         Z14159
   #authors       Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
                    Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
                    Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
                    J.C.
   #submission    submitted to the EMBL Data Library, November 1997
   #description   Arabidopsis thaliana chromosome II BAC T20D16 genomic
                    sequence.
   #accession     T00502
      ##status         translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-773 ##label ROU
      ##cross-references EMBL:AC002391; NID:g2642427; PID:g2642433
      ##experimental_source cultivar Columbia
GENETICS
   #map_position 2
   #introns       545/1
   #note          T20D16.7
SUMMARY           #length 773  #molecular-weight 84148  #checksum 123

  Query Match           79.4%;  Score 50;  DB 2;  Length 773;
  Best Local Similarity  77.8%;  Pred. No. 9.80e+00;
  Matches     7;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    343 AGIGILALI 351
          ||||||::|
Qy      2 AGIGILTVI 10


RESULT    6
ENTRY             T15544     #type complete
TITLE             hypothetical protein C18A3.2 - Caenorhabditis elegans
ORGANISM          #formal_name Caenorhabditis elegans
DATE              20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                    20-Sep-1999
ACCESSIONS        T15544
REFERENCE         Z18367
   #authors       Hallsworth, K.
   #submission    submitted to the EMBL Data Library, June 1995
   #description   The sequence of C. elegans cosmid C18A3.
   #accession     T15544
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type DNA
      ##residues       1-295 ##label HAL
      ##cross-references EMBL:U28944; NID:g861346; PID:g861350;
                         PIDN:AAA68371.1; CESP:C18A3.2
      ##experimental_source strain Bristol N2
GENETICS
   #gene          CESP:C18A3.2
   #introns       69/3; 93/3; 198/3; 250/3
SUMMARY           #length 295  #molecular-weight 32258  #checksum 8445

  Query Match           77.8%;  Score 49;  DB 2;  Length 295;
  Best Local Similarity  70.0%;  Pred. No. 1.54e+01;
  Matches     7;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    278 AAGIGIFVII 287
          |||||| :|
Qy      1 AAGIGILTVI 10


RESULT    7
ENTRY             B69783     #type complete
TITLE             transporter homolog ydgH - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    24-Sep-1998
ACCESSIONS        B69783
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #cross-references MUID:98044033
   #accession     B69783
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-885 ##label KUN
      ##cross-references GB:Z99106; GB:Z99107; GB:AL009126; NID:g2632866;
                         PID:e1182544; PID:g2632878; NID:g2632653;
                         PID:e1182531; PID:g2632865
      ##experimental_source strain 168
GENETICS
   #gene          ydgH
SUMMARY           #length 885  #molecular-weight 95488  #checksum 3557

  Query Match           77.8%;  Score 49;  DB 2;  Length 885;
  Best Local Similarity  60.0%;  Pred. No. 1.54e+01;
  Matches     6;  Conservative    2;  Mismatches   2;  Indels   0;  Gaps   0;

Db    305 AVGVGILMII 314
          | |:||| :|
Qy      1 AAGIGILTVI 10
```

```
RESULT    8
ENTRY             S48276      #type complete
TITLE             YSA1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES   protein YBR0907; protein YBR111c
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
                    10-Sep-1999
ACCESSIONS        S48276; S45979; S25364; S44691
REFERENCE         S48255
   #authors       Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann,
                    H.
   #journal       Yeast (1994) 10:1363-1381
   #title         Analysis of a 70 kb region on the right arm of yeast
                    chromosome II.
   #cross-references MUID:95208357
   #accession     S48276
      ##status          nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues       1-231 ##label MAN
      ##cross-references EMBL:X78993; NID:g476045; PID:g476067
REFERENCE         S45927
   #authors       Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
   #submission    submitted to the Protein Sequence Database, August 1994
   #accession     S45979
      ##molecule_type DNA
      ##residues       1-231 ##label FE2
      ##cross-references EMBL:Z35980; NID:g536465; PID:g536466; MIPS:YBR111c
REFERENCE         S25364
   #authors       Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann,
                    H.
   #journal       Yeast (1992) 8:397-408
   #title         Molecular analysis of yeast chromosome II between CMD1 and
                    LYS2: the excision repair gene RAD16 located in this region
                    belongs to a novel group of double-finger proteins.
   #cross-references MUID:92327848
   #accession     S25364
      ##molecule_type DNA
      ##residues       1-47 ##label MAW
      ##cross-references EMBL:X66247; NID:g3548; PID:g3549
GENETICS
   #gene          SGD:YSA1
      ##cross-references SGD:S0000315; MIPS:YBR111c
   #map_position 2R
CLASSIFICATION    #superfamily yffH protein; mutT domain homology
FEATURE
   111-145             #domain mutT domain homology #label MUTT
SUMMARY           #length 231  #molecular-weight 26087  #checksum 4809

  Query Match           76.2%;  Score 48;  DB 1;  Length 231;
  Best Local Similarity  75.0%;  Pred. No. 2.40e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     79 GIGILTIL 86
          ||||||::
Qy      3 GIGILTVI 10


RESULT    9
ENTRY             F69798      #type complete
TITLE             conserved hypothetical protein yetI - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    24-Sep-1998
ACCESSIONS        F69798
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #cross-references MUID:98044033
   #accession     F69798
      ##status          preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-361 ##label KUN
      ##cross-references GB:Z99107; GB:AL009126; NID:g2632866; PID:e1182696;
                         PID:g2633030
      ##experimental_source strain 168
GENETICS
   #gene          yetI
SUMMARY           #length 361  #molecular-weight 41966  #checksum 4304

  Query Match           76.2%;  Score 48;  DB 2;  Length 361;
  Best Local Similarity  70.0%;  Pred. No. 2.40e+01;
  Matches     7;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db    341 ADGIGILPLI 350
          | |||||  :|
Qy      1 AAGIGILTVI 10


RESULT   10
ENTRY             C70574      #type complete
TITLE             probable aroP2 protein - Mycobacterium tuberculosis (strain
                    H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    20-Aug-1999
ACCESSIONS        C70574
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537-544
```

```
   #title        Deciphering the biology of Mycobacterium tuberculosis from
                   the complete genome sequence.
   #cross-references MUID:98295987
   #accession    C70574
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-487 ##label COL
      ##cross-references GB:Z95324; GB:AL123456; NID:g3261760;
                        PIDN:CAB08578.1; PID:e315461; PID:g2094825
      ##experimental_source strain H37Rv
GENETICS
   #gene         aroP2
CLASSIFICATION   #superfamily arginine permease
SUMMARY          #length 487  #molecular-weight 52194  #checksum 4052

  Query Match           76.2%;  Score 48;  DB 2;  Length 487;
  Best Local Similarity  60.0%;  Pred. No. 2.40e+01;
  Matches     6;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    351 TAGIGLLGII 360
          :||||:| :|
Qy      1 AAGIGILTVI 10


RESULT   11
ENTRY            F70682     #type complete
TITLE            probable membrane protein - Mycobacterium tuberculosis
                   (strain H37RV)
ORGANISM         #formal_name Mycobacterium tuberculosis
DATE             17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                   17-Jul-1998
ACCESSIONS       F70682
REFERENCE        A70500
   #authors      Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                   C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                   III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                   Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                   Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                   Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                   Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                   Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                   Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                   Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal      Nature (1998) 393:537-544
   #title        Deciphering the biology of Mycobacterium tuberculosis from
                   the complete genome sequence.
   #cross-references MUID:98295987
   #accession    F70682
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-667 ##label COL
      ##cross-references GB:Z81368; GB:AL123456; NID:g3261656; PID:e279647;
                        PID:g1655665
      ##experimental_source strain H37Rv
GENETICS
   #gene         Rv2395
SUMMARY          #length 667  #molecular-weight 68251  #checksum 9102

  Query Match           76.2%;  Score 48;  DB 2;  Length 667;
  Best Local Similarity  55.6%;  Pred. No. 2.40e+01;
  Matches     5;  Conservative    3;  Mismatches   1;  Indels   0;  Gaps   0;

Db    379 SGVGILVVV 387
          :|:||| |:
Qy      2 AGIGILTVI 10


RESULT   12
ENTRY            S72755     #type complete
TITLE            B1496_C2_163 protein - Mycobacterium leprae
ORGANISM         #formal_name Mycobacterium leprae
DATE             19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
                   09-Sep-1997
ACCESSIONS       S72755
REFERENCE        S72693
   #authors      Smith, D.R.; Robison, K.
   #submission   submitted to the EMBL Data Library, November 1993
   #description  Mycobacterium leprae cosmid B1496.
   #accession    S72755
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-91 ##label SMI
      ##cross-references EMBL:U00013; NID:g466868; PID:g466877
GENETICS
   #start_codon  GTG
SUMMARY          #length 91  #molecular-weight 9561  #checksum 2921

  Query Match           74.6%;  Score 47;  DB 2;  Length 91;
  Best Local Similarity  66.7%;  Pred. No. 3.71e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     56 AGIGVLSAI 64
          ||||:|: |
Qy      2 AGIGILTVI 10


RESULT   13
ENTRY            A33351     #type complete
TITLE            H+-transporting ATP synthase (EC 3.6.1.34) proteolipid chain
                   - Sulfolobus acidocaldarius
ORGANISM         #formal_name Sulfolobus acidocaldarius
DATE             20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change
                   22-Jun-1999
ACCESSIONS       A33351
REFERENCE        A33351
   #authors      Denda, K.; Konishi, J.; Oshima, T.; Date, T.; Yoshida, M.
   #journal      J. Biol. Chem. (1989) 264:7119-7121
   #title        A gene encoding the proteolipid subunit of Sulfolobus
                   acidocaldarius ATPase complex.
   #cross-references MUID:89214142
   #accession    A33351
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-101 ##label DEN
      ##cross-references GB:J04740; NID:g152922; PIDN:AAA72703.1; PID:g152925
CLASSIFICATION   #superfamily H+-transporting ATP synthase lipid-binding
                   protein
KEYWORDS         hydrolase
SUMMARY          #length 101  #molecular-weight 10362  #checksum 4300

  Query Match           74.6%;  Score 47;  DB 2;  Length 101;
  Best Local Similarity  87.5%;  Pred. No. 3.71e+01;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 AAGIGVLT 66
          |||||:||
Qy      1 AAGIGILT 8


RESULT   14
ENTRY            C70959     #type complete
TITLE            hypothetical protein Rv1382 - Mycobacterium tuberculosis
                   (strain H37RV)
ORGANISM         #formal_name Mycobacterium tuberculosis
DATE             17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                   26-Aug-1999
ACCESSIONS       C70959
REFERENCE        A70500
   #authors      Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                   C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                   III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                   Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                   Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                   Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
```

```
                  Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                  Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                  Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                  Taylor, K.; Whitehead, S.; Barrell, B.G.
  #journal        Nature (1998) 393:537-544
  #title          Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
  #cross-references MUID:98295987
  #accession    C70959
     ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
     ##molecule_type DNA
     ##residues      1-165 ##label COL
     ##cross-references GB:Z81011; GB:AL123456; NID:g3242274; PID:e275153;
                         PID:g1621264
     ##experimental_source strain H37Rv
GENETICS
   #gene           Rv1382
CLASSIFICATION    #superfamily Mycobacterium tuberculosis hypothetical protein
                    Rv1382
SUMMARY           #length 165  #molecular-weight 18189  #checksum 5780

  Query Match             74.6%;  Score 47;  DB 2;  Length 165;
  Best Local Similarity   75.0%;  Pred. No. 3.71e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     128 AGIGILAI 135
           ||||||::
Qy       2 AGIGILTV 9


RESULT   15
ENTRY             A69843     #type complete
TITLE             hypothetical protein yjbA - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    24-Sep-1998
ACCESSIONS        A69843
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #cross-references MUID:98044033
   #accession    A69843
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-250 ##label KUN
      ##cross-references GB:Z99110; GB:AL009126; NID:g2633472; PID:e1183161;
                         PID:g2633495
      ##experimental_source strain 168
GENETICS
   #gene           yjbA
SUMMARY           #length 250  #molecular-weight 30119  #checksum 5271

  Query Match             74.6%;  Score 47;  DB 2;  Length 250;
  Best Local Similarity   66.7%;  Pred. No. 3.71e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db      89 TDGIGILAV 97
           : |||||:|
Qy       1 AAGIGILTV 9


Search completed: Fri May  5 22:17:24 2000
Job time : 60 secs.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:15:32 2000; MasPar time 3.20 Seconds
74.117 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-18
Description: (1-10) from US09267439.pep
Perfect Score: 63
Sequence: 1 AAGIGILTVI 10

Scoring table: PAM 150
Gap 15

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 15.856; Variance 51.362; scale 0.309

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	63	100.0	10	1	W98934	Human leukocyte antige	4.21e+00
2	63	100.0	10	1	R84198	MART-1 melanoma antige	4.21e+00
3	63	100.0	10	1	W07381	MART-1 epitope recogni	4.21e+00
4	63	100.0	21	1	W00903	Human melanoma MART-1/	4.21e+00
5	63	100.0	118	1	W83134	Human tumour rejection	4.21e+00
6	63	100.0	118	1	R63158	Tumour rejection antig	4.21e+00
7	63	100.0	118	1	R84212	MART-1 melanoma antige	4.21e+00
8	56	88.9	9	1	Y01751	Exemplary antigenic pe	2.38e+01
9	56	88.9	9	1	Y00713	Tumour antigen booster	2.38e+01
10	56	88.9	9	1	Y10444	HLA Class I motif pept	2.38e+01
11	56	88.9	9	1	Y10567	HLA Class I motif pept	2.38e+01
12	56	88.9	9	1	Y10601	HLA Class I motif pept	2.38e+01
13	56	88.9	9	1	W54602	Peptide 1 from Melan-A	2.38e+01
14	56	88.9	9	1	W07379	MART-1 epitope recogni	2.38e+01
15	56	88.9	9	1	W77123	MART-1/MelanA syntheti	2.38e+01
16	56	88.9	9	1	W68380	Human MART1/MELAN-A pe	2.38e+01
17	56	88.9	9	1	W39430	Human immunogenic T ce	2.38e+01
18	56	88.9	9	1	R84196	MART-1 melanoma antige	2.38e+01
19	56	88.9	9	1	W98938	Human leukocyte antige	2.38e+01
20	56	88.9	9	1	W35512	MART-1/Melan-A protein	2.38e+01
21	56	88.9	9	1	W42523	Melan A/MART epitope (	2.38e+01
22	56	88.9	10	1	W98939	Human leukocyte antige	2.38e+01
23	56	88.9	10	1	Y01750	Exemplary antigenic pe	2.38e+01
24	56	88.9	10	1	W98922	Human leukocyte antige	2.38e+01
25	56	88.9	10	1	Y00712	Tumour antigen booster	2.38e+01
26	56	88.9	10	1	W32269	Tumour rejection antig	2.38e+01
27	56	88.9	10	1	W54809	Peptide 1 from Mart-1/	2.38e+01
28	56	88.9	10	1	W39447	Human HLA-A*0201 immun	2.38e+01
29	56	88.9	10	1	W98924	Human leukocyte antige	2.38e+01
30	56	88.9	10	1	W98923	Human leukocyte antige	2.38e+01
31	56	88.9	10	1	W22039	Antigenic MART-1 pepti	2.38e+01
32	56	88.9	10	1	R84197	MART-1 melanoma antige	2.38e+01
33	56	88.9	10	1	W07380	MART-1 epitope recogni	2.38e+01
34	56	88.9	12	1	W22038	Antigenic MART-1 pepti	2.38e+01
35	55	87.3	9	1	W68381	Human MART1/MELAN-A pe	3.03e+01
36	55	87.3	9	1	R84764	MART-1 melanoma antige	3.03e+01
37	53	84.1	9	1	W42524	Melan A/MART (residues	4.91e+01
38	53	84.1	9	1	W42531	Melan A/MART epitope (	4.91e+01
39	53	84.1	9	1	W42525	Melan A/MART epitope (	4.91e+01
40	53	84.1	10	1	W98936	Human leukocyte antige	4.91e+01
41	52	82.5	9	1	R84786	Modified MART-1 melano	6.24e+01
42	52	82.5	9	1	R84788	Modified MART-1 melano	6.24e+01
43	52	82.5	9	1	R84787	Modified MART-1 melano	6.24e+01
44	52	82.5	479	1	P50034	Sequence encoded by th	6.24e+01
45	52	82.5	479	1	P81013	Complete sequence of t	6.24e+01

ALIGNMENTS

RESULT 1
ID W98934 standard; peptide; 10 AA.
AC W98934;
DT 06-MAY-1999 (first entry)
DE Human leukocyte antigen A2 molecule binding peptide SEQ ID NO:3.
KW Human leukocyte antigen; HLA; HLA-A2 binding peptide; T cell;
KW cytolytic T cell; CTL.
OS Synthetic.
OS Homo sapiens.
PN WO9858951-A1.
PD 30-DEC-1998.
PF 18-JUN-1998; U12879.
PR 16-APR-1998; US-061388.
PR 23-JUN-1997; US-880963.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Cerottini J, Romero P, Valmori D;
DR WPI; 99-105609/09.
PT New decamer peptides which bind to HLA molecules - useful to
PT identify HLA-A2 positive cells and provoke T cells
PS Claim 13; Page 9; 45pp; English.
CC The present invention describes peptides which bind to an HLA-A2
CC molecule and have Val at the carboxy terminus, and either: (a) Ala, Tyr
CC or Phe at the amino terminus, and Ala at position 2 (P1); or (b) Glu at
CC the amino terminus, and Ala, Leu, or Met at positions 2 and 3, with the
CC proviso that Ala is not at both positions (P2). The present sequence
CC represents an HLA-A2 binding peptide. The peptides of the present
CC invention are used to identify HLA-A2 positive cells, provoke T cells,
CC and determine the presence of particular T cells including cytolytic
CC T cells (CTLs). They provide a better target than the prior art
CC CTL-stimulating peptide.
SQ Sequence 10 AA;

Query Match 100.0%; Score 63; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 4.21e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db      1 AAGIGILTVI 10
          ||||||||||
Qy      1 AAGIGILTVI 10
```

RESULT 2
ID R84198 standard; Peptide; 10 AA.
AC R84198;
DT 20-APR-1996 (first entry)
DE MART-1 melanoma antigen immunogenic peptide M10-4.
KW MART-1; M10-4; melanoma antigen recognised by T-cells; melanoma;

KW metastatic melanoma; tumour-associated antigen;
KW immunogenic peptide; diagnosis; prognosis; prophylaxis;
KW therapy; vaccine.
OS Synthetic.
PN WO9529193-A2.
PD 02-NOV-1995.
PF 21-APR-1995; U05063.
PR 22-APR-1994; US-231565.
PR 05-APR-1995; US-417174.
PA (USSH) US SEC DEPT HEALTH.
PI Kawakami Y, Rosenberg SA;
DR WPI; 95-382963/49.
PT DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT vectors, host cells and antibodies, used to detect, treat and
PT immunise animal against melanoma.
PS Claim 12; Page 122; 184pp; English.
CC Immunogenic peptide M10-4 is a derivative of peptide M9-2 (R84196)
CC which is based on the melanoma antigen (MART-1) (see R84212).
CC M9-2 may be modified to improve immunogenicity (see R84783-R84800)
CC and used in medicaments for the treatment or prevention (by
CC immunization) of melanoma. Antibodies against MART-1 and its
CC immunogenic peptides may be used in the detection and isolation of
CC MART-1 from a sample, the detection of which is indicative of a
CC disease state (melanoma or metastatic melanoma).
CC See also R84196.
SQ Sequence 10 AA;

Query Match 100.0%; Score 63; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 4.21e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTVI 10
||||||||||
Qy 1 AAGIGILTVI 10

RESULT 3
ID W07381 standard; Peptide; 10 AA.
AC W07381;
DT 28-JUL-1997 (first entry)
DE MART-1 epitope recognised by melanoma specific T cell receptor.
KW T cell; receptor; lymphocyte; alpha; beta chain; V; variable;
KW J; joining; D; diversity; gene segment; probe; detection;
KW recombination; melanoma; cancer; neoplasia; tumour; diagnosis;
KW MART; Melanoma Antigen Recognised by T lymphocyte.
OS Homo sapiens.
PN WO9630516-A1.
PD 03-OCT-1996.
PF 27-MAR-1996; U04143.
PR 27-MAR-1995; US-411098.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Hwu P, Nishimura M, Rosenberg SA;
DR WPI; 96-485449/48.
PT T cell receptor alpha and/or beta chains, and related nucleic acids
PT - useful in pharmaceutical compsns. to prevent or treat cancer,
PT partic. lung, melanoma, ovarian, colon, brain or kidney tumours
PS Example 3; Page 11; 125pp; English
CC W07378-W07381 are MART-1 epitopes, M9-1, M9-2, M10-3 and M10-4
CC respectively, that are recognised by melanoma specific T lymphocyte
CC receptors (TCRs). Melanoma-specific TCRs comprising an alpha and
CC beta chain were made. Nucleic acids from either of these chains can be
CC used as probes for the detection of expression of rearranged genes
CC encoding tumour-associated antigens. The nucleic acids may also be used
CC to create transgenic animals, useful as biological models to study cancer
CC and evaluate diagnostic and therapeutic methods for the treatment of
CC cancers, particularly melanomas. Antibodies (Abs) may be raised against
CC alpha and beta chain polypeptides and used to detect native or denatured
CC TCRs and/or alterations in expression levels of T cells carrying
CC melanoma-specific TCRs. Abs can also purify and enrich T cells carrying
CC the above receptors, which can then be administered therapeutically to
CC mammals. Anti-idiotype antibodies can be used to assess the level of a
CC specific T cell carrying these receptors in a mammal being treated using
CC these methods. Host cells and vectors carrying nucleic acid encoding
CC a TCR (or individual alpha or beta chain fragment) are useful in
CC pharmaceutical compositions to prevent or treat cancer in a mammal, e.g.
CC lung, melanoma, ovarian, colon, brain or kidney tumours.
SQ Sequence 10 AA;

Query Match 100.0%; Score 63; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 4.21e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTVI 10
||||||||||
Qy 1 AAGIGILTVI 10

RESULT 4
ID W00903 standard; Peptide; 21 AA.
AC W00903;
DT 23-MAY-1997 (first entry)
DE Human melanoma MART-1/Aa tumour associated antigen p27-47.
KW Adeno-associated virus; vector; liposome; transfection;
KW dendritic cell; melanoma; MART-1/Aa; adoptive immunotherapy;
KW tumour associated antigen.
OS Homo sapiens.
PN WO9703703-A1.
PD 06-FEB-1997.
PF 19-JUL-1996; U12012.
PR 21-JUL-1995; US-001312.
PR 01-NOV-1995; US-007184.
PR 01-DEC-1995; US-566286.
PA (RHON) RHONE POULENC RORER PHARM INC.
PI Lebkowski JS, Philip R;
DR WPI; 97-145208/13.
PT Adeno-associated virus:liposome complexes for transfecting dendritic
PT cells - for inducing immune response, useful for treating e.g.
PT neoplasia or infections
PS Example 5; Page 58; 134pp; English.
CC Tumour associated antigens (W13660-61, W00878-903) can be loaded
CC into dendritic cells and used to induce antitumour immunity.
CC Alternatively, the dendritic cells are transfected with adeno
CC associated virus plasmid DNA (which includes DNA encoding the
CC tumour associated antigen) complexed with cationic liposomes. The
CC amtigen loaded or transfected dendritic cells can be used to
CC generate tumour antigen-specific cytotoxic T lymphocytes for use in
CC adoptive immunotherapy in a patient having the corresponding
CC tumour. A suitable antigen comprises amino acids 27-47 (W00903)
CC of human melanoma MART-1/Aa.
SQ Sequence 21 AA;

Query Match 100.0%; Score 63; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.21e+00;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 AAGIGILTVI 10
||||||||||
Qy 1 AAGIGILTVI 10

RESULT 5
ID W83134 standard; Protein; 118 AA.
AC W83134;
DT 04-FEB-1999 (first entry)
DE Human tumour rejection antigen precursor.
KW Human; tumour rejection antigen precursor; human leukocyte antigen;
KW TRAP; HLA; cancer; melanoma.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 2
FT /note= "encoded by CGA"
FT Misc_difference 17
FT /note= "encoded by GAC"
PN US5837476-A.
PD 17-NOV-1998.
PF 16-JAN-1998; 007966.

```
PR   03-MAR-1995; US-398409.
PR   16-JAN-1998; US-007966.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon-Falleur T, Brichard V, De Plaen E, Traversari C,
PI   Van Pel A, Woelfelt;
DR   WPI; 99-043967/04.
DR   N-PSDB; V70150.
PT   Use of a tumour rejection antigen precursor - as a marker for
PT   diagnosing a disorder characterised by expression of a tumour
PT   rejection antigen precursor which is not tyrosinase
PS   Claim 1; Column 7-9; 11pp; English.
CC   A method has been developed for the diagnosis of a disorder which is
CC   characterised by the expression of a tumour rejection antigen precursor
CC   (TRAP) which is not tyrosinase, and which is processed to a TRA which
CC   forms a complex with an HLA-A2 molecule. The present sequence represents
CC   the TRAP for use in the present invention. The method comprises
CC   contacting a sample from a subject with an agent specific for the
CC   complex and determining the interaction between the complex and the
CC   agent as a determination of the disorder. TRAP can be used for the
CC   diagnosis and treatment of disorders characterised by the expression
CC   of the TRAP molecules such as cancers, particularly melanoma.
SQ   Sequence   118 AA;

  Query Match          100.0%;  Score 63;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 4.21e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 AAGIGILTVI 36
          ||||||||||
Qy      1 AAGIGILTVI 10


RESULT    6
ID   R63158 standard; Protein; 118 AA.
AC   R63158;
DT   26-MAY-1995  (first entry)
DE   Tumour rejection antigen precursor.
KW   Tumour rejection antigen; precursor; HLA-A2 molecule; tyrosinase;
KW   isolation; melanoma; cell line; LB-39-MEL; diagnosis; vaccine;
KW   therapy.
OS   Homo sapiens.
PN   WO9421126-A.
PD   29-SEP-1994.
PF   09-MAR-1994; U02487.
PR   18-MAR-1993; US-032978.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon-Falleur T,  Brichard V,  De Plaen E,  Traversari C;
PI   Van Pel A, Wolfel T;
DR   WPI; 94-316544/39.
DR   N-PSDB; Q76370.
PT   Nucleic acid coding for a tumour rejection antigen precursor - is
PT   used for developing prods. for diagnosis or treatment of expression
PT   related disorders, partic. melanoma
PS   Claim 5; Page 14; 26pp; English.
CC   This sequence represents the tumour rejection antigen precursor which is
CC   processed to a tumour rejection antigen presented by HLA-A2 molecules.
CC   The tumour rejection antigen is not related to tyrosinase.  The cDNA
CC   encoding this sequence was isolated from the melanoma cell line,
CC   LB-39-MEL.  The tumour rejection antigen may be used for diagnosis or
CC   in vaccines or for therapy of disorders characterised by the expression
CC   of the tumour rejection antigen precursor, particularly melanoma.
SQ   Sequence   118 AA;

  Query Match          100.0%;  Score 63;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 4.21e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 AAGIGILTVI 36
          ||||||||||
Qy      1 AAGIGILTVI 10


RESULT    7
ID   R84212 standard; Protein; 118 AA.
AC   R84212;
DT   20-APR-1996  (first entry)
DE   MART-1 melanoma antigen.
KW   MART-1; melanoma antigen recognised by T-cell; melanoma;
KW   metastatic melanoma; tumour-associated antigen; immunogen;
KW   diagnosis; prognosis; prophylaxis; therapy; vaccine.
OS   Mammalian.
FH   Key             Location/Qualifiers
FT   region          27. .47
FT                   /note= "hydrophobic region"
PN   WO9529193-A2.
PD   02-NOV-1995.
PF   21-APR-1995; U05063.
PR   22-APR-1994; US-231565.
PR   05-APR-1995; US-417174.
PA   (USSH ) US SEC DEPT HEALTH.
PI   Kawakami Y,  Rosenberg SA;
DR   WPI; 95-382963/49.
DR   N-PSDB; T02714.
PT   DNA encoding melanoma antigens recognised by T-lymphocytes - also
PT   vectors, host cells and antibodies, used to detect, treat and
PT   immunise animal against melanoma.
PS   Claim 11; Page 117; 184pp; English.
CC   The melanoma antigen (MART-1) is produced by recombinant DNA
CC   methods, i.e. preferably using a baculovirus vector for expression
CC   in insect cell cultures.  MART-1 protein is a source of immunogenic
CC   peptides (see R84196 for peptide M9-2) which are optionally modified
CC   (see R84783-R84800) and used in medicaments for the treatment or
CC   prevention (by immunization) of melanoma.  Antibodies against MART-1
CC   and its immunogenic peptides may be used in the detection and
CC   isolation of MART-1 from a sample, the detection of which is
CC   indicative of a disease state (melanoma or metastatic melanoma).
SQ   Sequence   118 AA;

  Query Match          100.0%;  Score 63;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 4.21e+00;
  Matches    10;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 AAGIGILTVI 36
          ||||||||||
Qy      1 AAGIGILTVI 10


RESULT    8
ID   Y01751 standard; Peptide; 9 AA.
AC   Y01751;
DT   25-JUN-1999  (first entry)
DE   Exemplary antigenic peptide derived from Melan-A(MART-1).
KW   MAGE-3; tumour associated gene; human leucocyte antigen Class II;
KW   autologous CD4+ cell; MAGE-3 related disease; cancer; melanoma;
KW   osteosarcoma; leukemia; carcinoma.
OS   Homo sapiens.
PN   WO9914326-A1.
PD   25-MAR-1999.
PF   04-SEP-1998; U18601.
PR   12-SEP-1997; US-928615.
PA   (LUDW-) LUDWIG INST CANCER RES.
PA   (UYVR-) UNIV VRIJE BRUSSEL.
PI   Boon-Falleur T, Chaux P, Corthals J, Heirman C,
PI   Luiten R, Stroobant V, Thielemans K, Van Der Bruggen P;
DR   WPI; 99-244031/20.
PT   Isolated peptides that bind to human leucocyte antigen class II
PT   molecules
PS   Disclosure; Page 29; 88pp; English.
CC   The present sequence represents an exemplary tumour associated peptide
CC   antigen. The specification describes a MAGE-3 tumour associated gene.
CC   Peptides (Y01721-25) that bind human leucocyte antigen (HLA) Class II
CC   molecules can be derived from the MAGE-3 protein. These peptides and
CC   autologous CD4+ cells that bind to a complex of MAGE-3 peptide
CC   and HLA Class II, are used to treat MAGE-3 related diseases,
CC   particularly cancers (e.g. melanoma, osteosarcoma, leukemia and
CC   various forms of carcinoma). The peptides are also used to produce
```

```
CC   specific antibodies. Detection of of the peptides, e.g. in binding
CC   assays, particularly with antibodies, is used for diagnosis of such
CC   diseases.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT    9
ID   Y00713 standard; peptide; 9 AA.
AC   Y00713;
DT   12-MAY-1999  (first entry)
DE   Tumour antigen booster peptide Melan-AMART-1 HLA-A2 #2.
KW   Tumour antigen; booster peptide; immune response modulation; allergy;
KW   immune response enhancer; tumour cell; tumour rejection antigen;
KW   leukocyte antigen-presenting molecule; autoimmune disease;
KW   allograft rejection.
OS   Homo sapiens.
PN   WO9858956-A2.
PD   30-DEC-1998.
PF   19-JUN-1998; U12894.
PR   23-JUN-1997; US-880979.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon-Falleur T, Uyttenhove C, Warnier G;
DR   WPI; 99-105612/09.
PT   Immunization methods using viruses expressing antigen for priming
PT   and booster immunizations - useful for modulating immune responses
PT   against antigen, e.g. enhancing immune response against tumour cells
PT   expressing tumour rejection antigens
PS   Disclosure; Page 10; 33pp; English.
CC   This sequence represents a tumour antigen booster peptide that can be
CC   used in the method of the invention. The method is for for modulating an
CC   immune response in a mammal against an antigen, and comprises:
CC   (A) inducing an immune response by: (i) administering a virus containing
CC   a nucleic acid molecule encoding the antigen or its precursor to generate
CC   an immune response; and (ii) administering at least one booster dose
CC   comprising a peptide including the antigen, in an adjuvant, in a combined
CC   amount effective to enhance the initial immune response; or
CC   (B) reducing an immune response as defined for (A) but using a
CC   non-adjuvant with the peptide which includes the antigen, in an amount
CC   effective to reduce the initial immune response. Method (A) is used to
CC   enhance the immune response against tumour cells expressing tumour
CC   rejection antigens, and against pathogens in subjects having human
CC   leukocyte antigen-presenting molecules. Method (B) is used to reduce the
CC   immune response in allergy, autoimmune disease, and allograft rejection.
CC   Method (A) provides an immunisation method which, unlike prior art, is
CC   not limited by the host immune response against viral vectors.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   10
ID   Y10444 standard; Peptide; 9 AA.
AC   Y10444;
DT   12-MAY-1999  (first entry)
DE   HLA Class I motif peptide SEQ ID NO:374.
KW   Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW   immunisation; tumour; infectious disease; immunotherapy; cancer;
KW   malignant melanoma; viral disease; hepatitis; AIDS.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9902183-A2.
PD   21-JAN-1999.
PF   10-JUL-1998; U14289.
PR   10-DEC-1997; US-988320.
PR   10-JUL-1997; CA-209815.
PA   (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI   Kuendig TM, Simard JJL;
DR   WPI; 99-120514/10.
PT   Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT   of antigen in the lymphatic system of a mammal so as to provide a
PT   sustained CTL response, used to treat, e.g. AIDS
PS   Disclosure; Page 40; 199pp; English.
CC   The present invention describes a method of inducing and/or sustaining
CC   an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC   method comprises: (a) delivering an antigen to the mammal at a level to
CC   induce an immunological CTL response in the mammal; and (b) maintaining
CC   the level of the antigen in the mammal's lymphatic system to maintain
CC   the immunologic CTL response. The method can be used for the delivery of
CC   e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC   an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC   gene antigen, or a viral antigen. They can be used for the treatment of
CC   disease such as cancer, e.g. malignant melanoma or infectious disease,
CC   e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC   to the lymphatic system provides for potent CTL stimulation that takes
CC   place in the milieu of the lymphoid organ, and it sustains stimulation
CC   that is necessary to keep CTL active, cytotoxic and recirculating
CC   through the body. Y10071 to Y10639 represent examples of peptide
CC   antigens given in the present invention.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   11
ID   Y10567 standard; Peptide; 9 AA.
AC   Y10567;
DT   12-MAY-1999  (first entry)
DE   HLA Class I motif peptide SEQ ID NO:497.
KW   Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW   immunisation; tumour; infectious disease; immunotherapy; cancer;
KW   malignant melanoma; viral disease; hepatitis; AIDS.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9902183-A2.
PD   21-JAN-1999.
PF   10-JUL-1998; U14289.
PR   10-DEC-1997; US-988320.
PR   10-JUL-1997; CA-209815.
PA   (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI   Kuendig TM, Simard JJL;
DR   WPI; 99-120514/10.
PT   Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT   of antigen in the lymphatic system of a mammal so as to provide a
PT   sustained CTL response, used to treat, e.g. AIDS
PS   Disclosure; Page 47; 199pp; English.
CC   The present invention describes a method of inducing and/or sustaining
CC   an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC   method comprises: (a) delivering an antigen to the mammal at a level to
CC   induce an immunological CTL response in the mammal; and (b) maintaining
CC   the level of the antigen in the mammal's lymphatic system to maintain
CC   the immunologic CTL response. The method can be used for the delivery of
CC   e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC   an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC   gene antigen, or a viral antigen. They can be used for the treatment of
CC   disease such as cancer, e.g. malignant melanoma or infectious disease,
CC   e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
```

```
CC   to the lymphatic system provides for potent CTL stimulation that takes
CC   place in the milieu of the lymphoid organ, and it sustains stimulation
CC   that is necessary to keep CTL active, cytotoxic and recirculating
CC   through the body. Y10071 to Y10639 represent examples of peptide
CC   antigens given in the present invention.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   12
ID   Y10601 standard; Peptide; 9 AA.
AC   Y10601;
DT   12-MAY-1999  (first entry)
DE   HLA Class I motif peptide SEQ ID NO:531.
KW   Cytotoxic T-lymphocyte response; CTL; antigen; lymphatic system;
KW   immunisation; tumour; infectious disease; immunotherapy; cancer;
KW   malignant melanoma; viral disease; hepatitis; AIDS.
OS   Synthetic.
OS   Homo sapiens.
PN   WO9902183-A2.
PD   21-JAN-1999.
PF   10-JUL-1998; U14289.
PR   10-DEC-1997; US-988320.
PR   10-JUL-1997; CA-209815.
PA   (CTLI-) CTL IMMUNOTHERAPIES CORP.
PI   Kuendig TM, Simard JJL;
DR   WPI; 99-120514/10.
PT   Inducing a cytotoxic T lymphocyte response - by maintaining a level
PT   of antigen in the lymphatic system of a mammal so as to provide a
PT   sustained CTL response, used to treat, e.g. AIDS
PS   Disclosure; Page 49; 199pp; English.
CC   The present invention describes a method of inducing and/or sustaining
CC   an immunological cytotoxic T lymphocyte (CTL) response in a mammal. The
CC   method comprises: (a) delivering an antigen to the mammal at a level to
CC   induce an immunological CTL response in the mammal; and (b) maintaining
CC   the level of the antigen in the mammal's lymphatic system to maintain
CC   the immunologic CTL response. The method can be used for the delivery of
CC   e.g. a differentiation antigen, a tumour-specific multilineage antigen,
CC   an embryonic antigen, an oncogene antigen, a mutated tumour-suppressor
CC   gene antigen, or a viral antigen. They can be used for the treatment of
CC   disease such as cancer, e.g. malignant melanoma or infectious disease,
CC   e.g. viral disease such as hepatitis or AIDS. Sustained antigen delivery
CC   to the lymphatic system provides for potent CTL stimulation that takes
CC   place in the milieu of the lymphoid organ, and it sustains stimulation
CC   that is necessary to keep CTL active, cytotoxic and recirculating
CC   through the body. Y10071 to Y10639 represent examples of peptide
CC   antigens given in the present invention.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   13
ID   W54602 standard; peptide; 9 AA.
AC   W54602;
DT   25-SEP-1998  (first entry)
DE   Peptide 1 from Melan-A/Mart-1.
KW   Mannose; antigen; antigen-presenting cell; mannosylated peptide; T cell;
KW   vaccine; treatment.
OS   Synthetic.
PN   WO9813378-A1.
PD   02-APR-1998.
PF   25-SEP-1997; NL0536.
PR   26-SEP-1996; EP-202701.
PA   (UYLE-) RIJKSUNIV LEIDEN.
PI   Drijfhout JW, Koning F;
DR   WPI; 98-230631/20.
PT   Increasing uptake and presentation of antigen(s) - by adding mannose
PT   residue(s) to antigen for increasing T cell response, useful in,
PT   e.g. vaccines against viral infection(s)
PS   Disclosure; Page 24; 47pp; English.
CC   The peptides W54559-W54809 are examples of peptides to which at least 1
CC   (preferably 2) mannose can be attached to increase their uptake as
CC   antigens by antigen-presenting cells.  Uptake of agonist mannosylated
CC   peptides will increase the T cell response, whereas uptake of antagonist
CC   peptides blocks the T cell response.  Blocking binding of immunogenic
CC   autoantigens can be used in treatment of type I diabetes, rheumatoid
CC   arthritis, graft rejection etc., also to induce T-cell non-
CC   responsiveness.  Vaccines containing mannosylated antigen are used to
CC   prevent or treat infections by, e.g. bacteria, viruses, fungi, helminths
CC   and parasites.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   14
ID   W07379 standard; Peptide; 9 AA.
AC   W07379;
DT   28-JUL-1997  (first entry)
DE   MART-1 epitope recognised by melanoma specific T cell receptor.
KW   T cell; receptor; lymphocyte; alpha; beta chain; V; variable;
KW   J; joining; D; diversity; gene segment; probe; detection;
KW   recombination; melanoma; cancer; neoplasia; tumour; diagnosis;
KW   MART; Melanoma Antigen Recognised by T lymphocyte.
OS   Homo sapiens.
PN   WO9630516-A1.
PD   03-OCT-1996.
PF   27-MAR-1996; U04143.
PR   27-MAR-1995; US-411098.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Hwu P,  Nishimura M,  Rosenberg SA;
DR   WPI; 96-485449/48.
PT   T cell receptor alpha and/or beta chains, and related nucleic acids
PT   - useful in pharmaceutical compsns. to prevent or treat cancer,
PT   partic. lung, melanoma, ovarian, colon, brain or kidney tumours
PS   Example 3; Page 11; 125pp; English
CC   W07378-W07381 are MART-1 epitopes, M9-1, M9-2, M10-3 and M10-4
CC   respectively, that are recognised by melanoma specific T lymphocyte
CC   receptors (TCRs). Melanoma-specific TCRs comprising an alpha and
CC   beta chain were made. Nucleic acids from either of these chains can be
CC   used as probes for the detection of expression of rearranged genes
CC   encoding tumour-associated antigens. The nucleic acids may also be used
CC   to create transgenic animals, useful as biological models to study cancer
CC   and evaluate diagnostic and therapeutic methods for the treatment of
CC   cancers, particularly melanomas. Antibodies (Abs) may be raised against
CC   alpha and beta chain polypeptides and used to detect native or denatured
CC   TCRs and/or alterations in expression levels of T cells carrying
CC   melanoma-specific TCRs. Abs can also purify and enrich T cells carrying
CC   the above receptors, which can then be administered therapeutically to
CC   mammals. Anti-idiotype antibodies can be used to assess the level of a
CC   specific T cell carrying these receptors in a mammal being treated using
CC   these methods. Host cells and vectors carrying nucleic acid encoding
CC   a TCR (or individual alpha or beta chain fragment) are useful in
CC   pharmaceutical compositions to prevent or treat cancer in a mammal, e.g.
CC   lung, melanoma, ovarian, colon, brain or kidney tumours.
SQ   Sequence   9 AA;
```

```
  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


RESULT   15
ID   W77123 standard; peptide; 9 AA.
AC   W77123;
DT   16-NOV-1998  (first entry)
DE   MART-1/MelanA synthetic peptide epitope 1.
KW   Tyrosinase; tyrosinase cytotoxic lymphocyte response;
KW   cytotoxic T lymphocyte; cysteine-depleted; melanoma.
OS   Synthetic.
PN   WO9833810-A2.
PD   06-AUG-1998.
PF   29-JAN-1998; U01592.
PR   30-JAN-1997; US-037781.
PA   (UYVI-) UNIV VIRGINIA PATENT FOUND.
PI   Engelhard VH, Hunt DF, Kittlesen D, Slingluff CL;
DR   WPI; 98-437388/37.
PT   Disease specific immunogen  - comprises disease specific cytotoxic T
PT   lymphocyte epitope used to elicit melanoma specific CTL response
PS   Disclosure; Page 27; 93pp; English.
CC   The peptide epitope W77119-W77138 were created for human tumour-specific
CC   cytotoxic T lymphocyte response.  These peptides are are cysteine-
CC   depleted mutants of a native disease-specific CTL epitope.  The cysteine-
CC   depleted CTL epitopes elicit a stronger or more specific CTL response
CC   than the native epitope.  The epitopes can be used in a disease-specific
CC   immunogen to protect a mammal against disease in particular melanomas.
CC   The peptides may also be used to screen a sample for the presence of
CC   an antigen with the same epitope, or with a different cross-reactive
CC   epitope.
SQ   Sequence   9 AA;

  Query Match           88.9%;  Score 56;  DB 1;  Length 9;
  Best Local Similarity 100.0%;  Pred. No. 2.38e+01;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 AAGIGILTV 9
          |||||||||
Qy      1 AAGIGILTV 9


Search completed: Fri May  5 22:16:06 2000
Job time : 34 secs.
```

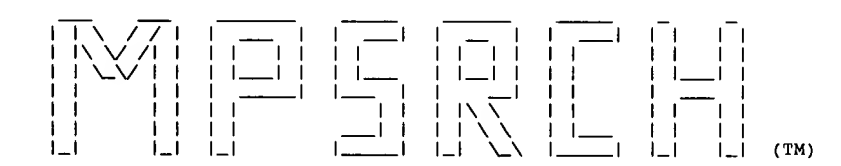
```
DATE              30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
                    22-Jun-1999
ACCESSIONS        S13592; B33358
REFERENCE         S13592
   #authors       Gardner, P.D.
   #journal       Nucleic Acids Res. (1990) 18:6714
   #title         Nucleotide sequence of the epsilon-subunit of the mouse
                    muscle nicotinic acetylcholine receptor.
   #cross-references MUID:91067487
   #accession     S13592
      ##molecule_type mRNA
      ##residues      1-493 ##label GAR
      ##cross-references EMBL:X55718; NID:g53160; PIDN:CAA39251.1; PID:g53161
REFERENCE         A33358
   #authors       Buonanno, A.; Mudd, J.; Merlie, J.P.
   #journal       J. Biol. Chem. (1989) 264:7611-7616
   #title         Isolation and characterization of the beta-and
                    epsilon-subunit genes of mouse muscle acetylcholine
                    receptor.
   #cross-references MUID:89214211
   #accession     B33358
      ##molecule_type DNA
      ##residues      1-493 ##label BUO
      ##cross-references GB:J04698; NID:g191599; PIDN:AAA37153.1; PID:g387086
CLASSIFICATION    #superfamily acetylcholine receptor
KEYWORDS          glycoprotein; ion channel; muscle; neurotransmitter receptor;
                    postsynaptic membrane; transmembrane protein

FEATURE
   1-20               #domain signal sequence #status predicted #label SIG\
   21-493             #product nicotinic acetylcholine receptor epsilon chain
                        #status predicted #label MAT\
   21-239             #domain extracellular #status predicted #label EXT\
   240-266            #domain transmembrane #status predicted #label TM1\
   273-291            #domain transmembrane #status predicted #label TM2\
   307-328            #domain transmembrane #status predicted #label TM3\
   329-456            #domain intracellular #status predicted #label INT\
   457-479            #domain transmembrane #status predicted #label TM4\
   86,161,327         #binding_site carbohydrate (Asn) (covalent) #status
                        predicted\
   148-162            #disulfide_bonds #status predicted
SUMMARY           #length 493  #molecular-weight 54914  #checksum 1794

  Query Match           76.8%;  Score 43;  DB 1;  Length 493;
  Best Local Similarity  75.0%;  Pred. No. 6.64e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db      93 AGVGILRV 100
           ||:||| |
Qy       2 AGIGILTV 9


Search completed: Fri May  5 22:00:34 2000
Job time : 62 secs.
```

```
   #authors      Gosalbes, M.J.; Perez-Gonzalez, J.A.; Gonzalez, R.; Navarro,
                   A.
   #journal      J. Bacteriol. (1991) 173:7705-7710
   #title        Two beta-glycanase genes are clustered in Bacillus polymyxa:
                   molecular cloning, expression, and sequence analysis of
                   genes encoding a xylanase and an endo-beta-(1,3)-(1,
                   4)-glucanase.
   #cross-references MUID:92041687
   #accession    S19011
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-635 ##label GOS
      ##cross-references EMBL:X57094; NID:g48815; PID:g48816
      ##note          the authors translated the codon GAA for residue 78 as
                       Gly, CCT for residue 272 as Thr, ATC for residue 412
                       as Gln, and ATC for residue 478 as Tyr
FUNCTION
   #description  catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in
                   xylans
   #pathway      xylan degradation
CLASSIFICATION   #superfamily Clostridium xylanase A repeat homology
KEYWORDS         glycosidase; hydrolase; polysaccharide degradation
FEATURE
   408-502            #domain Clostridium xylanase A repeat homology #label
                        CXA
SUMMARY          #length 635  #molecular-weight 67914  #checksum 2077

  Query Match           78.6%;  Score 44;  DB 2;  Length 635;
  Best Local Similarity  75.0%;  Pred. No. 4.19e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    149 GAGIGVLT 156
          :||||:||
Qy      1 AAGIGILT 8


RESULT   13
ENTRY            S01903     #type complete
TITLE            H+-transporting ATP synthase (EC 3.6.1.34) epsilon chain -
                   Arabidopsis thaliana chloroplast
ORGANISM         #formal_name chloroplast Arabidopsis thaliana #common_name
                   mouse-ear cress
DATE             01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change
                   22-Jun-1999
ACCESSIONS       S01903
REFERENCE        S01903
   #authors      Chen, H.C.; Wintz, H.; Weil, J.H.; Pillay, D.T.N.
   #journal      Nucleic Acids Res. (1988) 16:10372
   #title        Nucleotide sequence of chloroplast CF1-ATPase epsilon-subunit
                   and elongator tRNA(Met) genes from Arabidopsis thaliana.
   #cross-references MUID:89057486
   #accession    S01903
      ##molecule_type DNA
      ##residues      1-132 ##label CHE
      ##cross-references EMBL:X12889; NID:g11332; PIDN:CAA31381.1; PID:g11334
GENETICS
   #gene         atpE
   #genome       chloroplast
CLASSIFICATION   #superfamily H+-transporting ATP synthase epsilon chain
KEYWORDS         ATP biosynthesis; chloroplast; hydrolase; membrane-associated
                   complex; thylakoid
SUMMARY          #length 132  #molecular-weight 14472  #checksum 1607

  Query Match           76.8%;  Score 43;  DB 2;  Length 132;
  Best Local Similarity  66.7%;  Pred. No. 6.64e+01;
  Matches     6;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     43 AVDIGILTI 51
          |  |||||:
Qy      1 AAGIGILTV 9


RESULT   14
ENTRY            D70073     #type complete
TITLE            metabolite transport protein homolog yxcC - Bacillus subtilis
ORGANISM         #formal_name Bacillus subtilis
DATE             05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                   24-Sep-1999
ACCESSIONS       D70073
REFERENCE        A69580
   #authors      Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                   Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                   Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                   A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                   Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                   Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                   Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                   Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                   Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                   M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                   S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                   Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                   C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                   Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                   Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                   Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                   Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                   Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                   Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                   M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                   M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                   V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                   A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                   Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                   Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                   Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                   Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                   B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                   Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                   Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                   Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                   Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                   Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                   K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                   Yoshikawa, H.; Danchin, A.
   #journal      Nature (1997) 390:249-256
   #title        The complete genome sequence of the Gram-positive bacterium
                   Bacillus subtilis.
   #cross-references MUID:98044033
   #accession    D70073
      ##status        preliminary; nucleic acid sequence not shown;
                       translation not shown
      ##molecule_type DNA
      ##residues      1-461 ##label KUN
      ##cross-references GB:Z99124; GB:AL009126; NID:g2636442;
                          PIDN:CAB16017.1; PID:e1184706; PID:g2636527
      ##experimental_source strain 168
GENETICS
   #gene         yxcC
CLASSIFICATION   #superfamily glucose transport protein
SUMMARY          #length 461  #molecular-weight 50140  #checksum 8642

  Query Match           76.8%;  Score 43;  DB 2;  Length 461;
  Best Local Similarity  85.7%;  Pred. No. 6.64e+01;
  Matches     6;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    285 GIGILNV 291
          ||||| |
Qy      3 GIGILTV 9


RESULT   15
ENTRY            ACMSE      #type complete
TITLE            nicotinic acetylcholine receptor epsilon chain precursor -
                   mouse
ORGANISM         #formal_name Mus musculus #common_name house mouse
```

```
                rats after partial hepatectomy and during sepsis.
   #cross-references MUID:96003917
   #accession    S59131
      ##status        preliminary
      ##molecule_type mRNA
      ##residues      1-420 ##label FUR
      ##cross-references EMBL:D43964; NID:g604901; PID:d1008487; PID:g604902
SUMMARY           #length 420  #molecular-weight 46496  #checksum 4868

  Query Match           80.4%;  Score 45;  DB 2;  Length 420;
  Best Local Similarity  55.6%;  Pred. No. 2.62e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    226 GPGVGILSV 234
          ::|:|||:|
Qy      1 AAGIGILTV 9


RESULT    9
ENTRY             D72106     #type complete
TITLE             hypothetical protein - Chlamydia pneumoniae (strain CWL029)
ORGANISM          #formal_name Chlamydia pneumoniae
DATE              23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
                    23-Apr-1999
ACCESSIONS        D72106
REFERENCE         A72000
   #authors       Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.;
                    Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
   #journal       Nature Genet. (1999) 21:385-389
   #title         Comparative genomes of Clamydia pneumoniae and C.
                    trachomatis.
   #cross-references MUID:99206606
   #accession    D72106
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-98 ##label ARN
      ##cross-references GB:AE001607; GB:AE001363; NID:g4376474; PID:g4376483
      ##experimental_source strain CWL029
GENETICS
   #gene          CPn0211
SUMMARY           #length 98  #molecular-weight 10280  #checksum 832

  Query Match           78.6%;  Score 44;  DB 2;  Length 98;
  Best Local Similarity  75.0%;  Pred. No. 4.19e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     54 AGVAILTV 61
          ||::||||
Qy      2 AGIGILTV 9


RESULT   10
ENTRY             C65176     #type complete
TITLE             glmU protein - Escherichia coli (strain K-12)
ORGANISM          #formal_name Escherichia coli
DATE              12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
                    29-Sep-1999
ACCESSIONS        C65176
REFERENCE         A64720
   #authors       Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
                    Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
                    Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
                    Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
                    Y.
   #journal       Science (1997) 277:1453-1462
   #title         The complete genome sequence of Escherichia coli K-12.
   #cross-references MUID:97426617
   #accession    C65176
      ##status        nucleic acid sequence not shown; translation not shown
      ##molecule_type DNA
      ##residues      1-456 ##label BLAT
      ##cross-references GB:AE000450; GB:U00096; NID:g1790166;
                        PIDN:AAC76753.1; PID:g1790168; UWGP:b3730
      ##experimental_source strain K-12, substrain MG1655
GENETICS
   #gene          glmU
CLASSIFICATION    #superfamily N-acetylglucosamine-1-phosphate
                    uridyltransferase
SUMMARY           #length 456  #molecular-weight 49190  #checksum 9400

  Query Match           78.6%;  Score 44;  DB 2;  Length 456;
  Best Local Similarity  75.0%;  Pred. No. 4.19e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    124 GGIGLLTV 131
          :|||:|||
Qy      2 AGIGILTV 9


RESULT   11
ENTRY             H69382     #type complete
TITLE             ABC transporter, ATP-binding protein homolog - Archaeoglobus
                    fulgidus
ORGANISM          #formal_name Archaeoglobus fulgidus
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    28-May-1999
ACCESSIONS        H69382
REFERENCE         A69250
   #authors       Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
                    K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
                    Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
                    D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
                    Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
                    Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
                    Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
                    Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
                    J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
                    T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
                    D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
                    Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
                    C.R.; Venter, J.C.
   #journal       Nature (1997) 390:364-370
   #title         The complete genome sequence of the hyperthermophilic,
                    sulfate-reducing archaeon Archaeoglobus fulgidus.
   #cross-references MUID:98049343
   #accession    H69382
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-620 ##label KLE
      ##cross-references GB:AE001029; GB:AE000782; NID:g2689352; PID:g2649523;
                        TIGR:AF1064
CLASSIFICATION    #superfamily ATP-binding cassette homology
KEYWORDS          ATP; P-loop
FEATURE
   428-612            #domain ATP-binding cassette homology #label ABC\
   445-452            #region nucleotide-binding motif A (P-loop)
SUMMARY           #length 620  #molecular-weight 71200  #checksum 5979

  Query Match           78.6%;  Score 44;  DB 2;  Length 620;
  Best Local Similarity  66.7%;  Pred. No. 4.19e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    283 ADGLGILAV 291
          | |:|||:|
Qy      1 AAGIGILTV 9


RESULT   12
ENTRY             S19011     #type complete
TITLE             endo-1,4-beta-xylanase (EC 3.2.1.8) - Bacillus polymyxa
ORGANISM          #formal_name Bacillus polymyxa
DATE              22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change
                    09-Apr-1998
ACCESSIONS        S19011
REFERENCE         S19011
```

```
          : ||||||:|
Qy      1 AAGIGILTV 9


RESULT    5
ENTRY             VGBEPB     #type complete
TITLE             glycoprotein gIII precursor - suid herpesvirus 1
ORGANISM          #formal_name suid herpesvirus 1
DATE              30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change
                    16-Jul-1999
ACCESSIONS        A26097
REFERENCE         A26097
   #authors       Robbins, A.K.; Watson, R.J.; Whealy, M.E.; Hays, W.W.;
                    Enquist, L.W.
   #journal       J. Virol. (1986) 58:339-347
   #title         Characterization of a Pseudorabies virus glycoprotein gene
                    with homology to herpes simplex virus type 1 and type 2
                    glycoprotein C.
   #cross-references MUID:86200375
   #accession     A26097
      ##molecule_type DNA
      ##residues       1-479 ##label ROB
      ##cross-references GB:M12778; NID:g334049; PIDN:AAA47464.1; PID:g334050
      ##experimental_source strain Becker
CLASSIFICATION    #superfamily herpesvirus glycoprotein F
KEYWORDS          glycoprotein
FEATURE
   1-22                #domain signal sequence #status predicted #label SIG\
   23-479              #product glycoprotein gIII #status predicted #label GPG\
   40,84,169,192,220,
   228,285,302         #binding_site carbohydrate (Asn) (covalent) #status
                         predicted
SUMMARY           #length 479  #molecular-weight 51206  #checksum 1630

  Query Match            83.9%;  Score 47;  DB 1;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 1.00e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      2 AGIGILTV 9


RESULT    6
ENTRY             S48276     #type complete
TITLE             YSA1 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES   protein YBR0907; protein YBR111c
ORGANISM          #formal_name Saccharomyces cerevisiae
DATE              10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
                    10-Sep-1999
ACCESSIONS        S48276; S45979; S25364; S44691
REFERENCE         S48255
   #authors       Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann,
                    H.
   #journal       Yeast (1994) 10:1363-1381
   #title         Analysis of a 70 kb region on the right arm of yeast
                    chromosome II.
   #cross-references MUID:95208357
   #accession     S48276
      ##status         nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues       1-231 ##label MAN
      ##cross-references EMBL:X78993; NID:g476045; PID:g476067
REFERENCE         S45927
   #authors       Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
   #submission    submitted to the Protein Sequence Database, August 1994
   #accession     S45979
      ##molecule_type DNA
      ##residues       1-231 ##label FE2
      ##cross-references EMBL:Z35980; NID:g536465; PID:g536466; MIPS:YBR111c
REFERENCE         S25364
   #authors       Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann,
                    H.
   #journal       Yeast (1992) 8:397-408
   #title         Molecular analysis of yeast chromosome II between CMD1 and
                    LYS2: the excision repair gene RAD16 located in this region
                    belongs to a novel group of double-finger proteins.
   #cross-references MUID:92327848
   #accession     S25364
      ##molecule_type DNA
      ##residues       1-47 ##label MAW
      ##cross-references EMBL:X66247; NID:g3548; PID:g3549
GENETICS
   #gene          SGD:YSA1
      ##cross-references SGD:S0000315; MIPS:YBR111c
   #map_position 2R
CLASSIFICATION    #superfamily yffH protein; mutT domain homology
FEATURE
   111-145             #domain mutT domain homology #label MUTT
SUMMARY           #length 231  #molecular-weight 26087  #checksum 4809

  Query Match            82.1%;  Score 46;  DB 1;  Length 231;
  Best Local Similarity  85.7%;  Pred. No. 1.63e+01;
  Matches     6;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     79 GIGILTI 85
          ||||||:
Qy      3 GIGILTV 9


RESULT    7
ENTRY             S62369     #type complete
TITLE             methylcobalamin--coenzyme M methyltransferase II -
                    Methanosarcina barkeri
ORGANISM          #formal_name Methanosarcina barkeri
DATE              19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change
                    17-Mar-1999
ACCESSIONS        S62369
REFERENCE         S62368
   #authors       Harms, U.; Thauer, R.K.
   #journal       Eur. J. Biochem. (1996) 235:653-659
   #title         Methylcobalamin:coenzyme M methyltransferase isoenzymes MtaA
                    and MtbA from Methanosarcina barkeri: cloning, sequencing
                    and differential transcription of the encoding genes, and
                    functional overexpression of the mtaA gene in Escherichia
                    coli.
   #cross-references MUID:96184544
   #accession     S62369
      ##status         preliminary; nucleic acid sequence not shown
      ##molecule_type DNA
      ##residues       1-339 ##label HAR
      ##cross-references EMBL:X91894; NID:g1107727; PID:e204100; PID:g1107728
SUMMARY           #length 339  #molecular-weight 36761  #checksum 6431

  Query Match            80.4%;  Score 45;  DB 2;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 2.62e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      2 AGIGILTV 9


RESULT    8
ENTRY             S59131     #type complete
TITLE             Kan-1 protein - rat
ORGANISM          #formal_name Rattus norvegicus #common_name Norway rat
DATE              15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
                    07-May-1999
ACCESSIONS        S59131
REFERENCE         S59131
   #authors       Furutani, M.; Arii, S.; Higashitsuji, H.; Mise, M.; Fukumoto,
                    M.; Takano, S.; Nakayama, H.; Imamura, M.; Fujita, J.
   #journal       Biochem. J. (1995) 311:203-208
   #title         Reduced expression of kan-1 (encoding putative bile
                    acid-CoA-amino acid N-acyltransferase) mRNA in livers of
```

```
Qy      1 AAGIGILTV 9


RESULT    2
ENTRY             A33351     #type complete
TITLE             H+-transporting ATP synthase (EC 3.6.1.34) proteolipid chain
                    - Sulfolobus acidocaldarius
ORGANISM          #formal_name Sulfolobus acidocaldarius
DATE              20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change
                    22-Jun-1999
ACCESSIONS        A33351
REFERENCE         A33351
   #authors       Denda, K.; Konishi, J.; Oshima, T.; Date, T.; Yoshida, M.
   #journal       J. Biol. Chem. (1989) 264:7119-7121
   #title         A gene encoding the proteolipid subunit of Sulfolobus
                    acidocaldarius ATPase complex.
   #cross-references MUID:89214142
   #accession     A33351
      ##status         preliminary
      ##molecule_type DNA
      ##residues       1-101 ##label DEN
      ##cross-references GB:J04740; NID:g152922; PIDN:AAA72703.1; PID:g152925
CLASSIFICATION    #superfamily H+-transporting ATP synthase lipid-binding
                    protein
KEYWORDS          hydrolase
SUMMARY           #length 101  #molecular-weight 10362  #checksum 4300

  Query Match           83.9%;  Score 47;  DB 2;  Length 101;
  Best Local Similarity  87.5%;  Pred. No. 1.00e+01;
  Matches     7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db      59 AAGIGVLT 66
           |||||:||
Qy      1 AAGIGILT 8


RESULT    3
ENTRY             C70959     #type complete
TITLE             hypothetical protein Rv1382 - Mycobacterium tuberculosis
                    (strain H37RV)
ORGANISM          #formal_name Mycobacterium tuberculosis
DATE              17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
                    26-Aug-1999
ACCESSIONS        C70959
REFERENCE         A70500
   #authors       Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
                    C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
                    III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.;
                    Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
                    Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
                    Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
                    Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
                    Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
                    Skelton, S.; Squares, S.; Sqares, R.; Sulston, J.E.;
                    Taylor, K.; Whitehead, S.; Barrell, B.G.
   #journal       Nature (1998) 393:537-544
   #title         Deciphering the biology of Mycobacterium tuberculosis from
                    the complete genome sequence.
   #cross-references MUID:98295987
   #accession     C70959
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-165 ##label COL
      ##cross-references GB:Z81011; GB:AL123456; NID:g3242274; PID:e275153;
                          PID:g1621264
      ##experimental_source strain H37Rv
GENETICS
   #gene          Rv1382
CLASSIFICATION    #superfamily Mycobacterium tuberculosis hypothetical protein
                    Rv1382
SUMMARY           #length 165  #molecular-weight 18189  #checksum 5780

  Query Match           83.9%;  Score 47;  DB 2;  Length 165;
  Best Local Similarity  75.0%;  Pred. No. 1.00e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     128 AGIGILAI 135
           ||||||::
Qy      2 AGIGILTV 9


RESULT    4
ENTRY             A69843     #type complete
TITLE             hypothetical protein yjbA - Bacillus subtilis
ORGANISM          #formal_name Bacillus subtilis
DATE              05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
                    24-Sep-1998
ACCESSIONS        A69843
REFERENCE         A69580
   #authors       Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
                    Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
                    Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
                    A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
                    Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
                    Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
                    Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
                    Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
                    Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
                    M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
                    S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
                    Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
                    C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
                    Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
                    Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
                    Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.;
                    Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
                    Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
                    Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
                    M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
                    M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
                    V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
                    A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
                    Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
                    Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
                    Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
                    Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
                    B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
                    Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
                    Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
                    Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
                    Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
                    Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
                    K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
                    Yoshikawa, H.; Danchin, A.
   #journal       Nature (1997) 390:249-256
   #title         The complete genome sequence of the Gram-positive bacterium
                    Bacillus subtilis.
   #cross-references MUID:98044033
   #accession     A69843
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-250 ##label KUN
      ##cross-references GB:Z99110; GB:AL009126; NID:g2633472; PID:e1183161;
                          PID:g2633495
      ##experimental_source strain 168
GENETICS
   #gene          yjbA
SUMMARY           #length 250  #molecular-weight 30119  #checksum 5271

  Query Match           83.9%;  Score 47;  DB 2;  Length 250;
  Best Local Similarity  66.7%;  Pred. No. 1.00e+01;
  Matches     6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db      89 TDGIGILAV 97
```



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 21:59:32 2000; MasPar time 4.97 Seconds
85.483 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-4
Description: (1-9) from US09267439.pep
Perfect Score: 56
Sequence: 1 AAGIGILTV 9

Scoring table: PAM 150
Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 22.221; Variance 25.904; scale 0.858

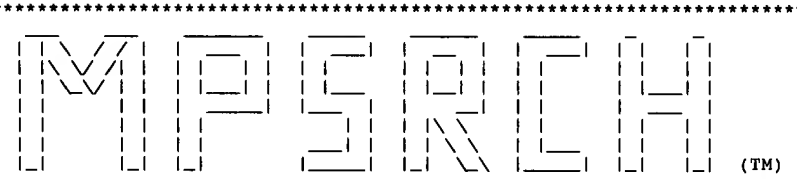
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	56	100.0	118	2	A55253	melanoma antigen MART	9.43e-02
2	47	83.9	101	2	A33351	H+-transporting ATP s	1.00e+01
3	47	83.9	165	2	C70959	hypothetical protein	1.00e+01
4	47	83.9	250	2	A69843	hypothetical protein	1.00e+01
5	47	83.9	479	1	VGBEPB	glycoprotein gIII pre	1.00e+01
6	46	82.1	231	1	S48276	YSA1 protein - yeast	1.63e+01
7	45	80.4	339	2	S62369	methylcobalamin--coen	2.62e+01
8	45	80.4	420	2	S59131	Kan-1 protein - rat	2.62e+01
9	44	78.6	98	2	D72106	hypothetical protein	4.19e+01
10	44	78.6	456	2	C65176	glmU protein - Escher	4.19e+01
11	44	78.6	620	2	H69382	ABC transporter, ATP-	4.19e+01
12	44	78.6	635	2	S19011	endo-1,4-beta-xylanas	4.19e+01
13	43	76.8	132	2	S01903	H+-transporting ATP s	6.64e+01
14	43	76.8	461	2	D70073	metabolite transport	6.64e+01
15	43	76.8	493	1	ACMSE	nicotinic acetylcholi	6.64e+01
16	43	76.8	509	2	H70597	probable membrane pro	6.64e+01
17	43	76.8	580	2	B70868	probable transferase	6.64e+01
18	43	76.8	593	1	A69655	two-component sensor	6.64e+01
19	43	76.8	611	2	JT0592	hypothetical protein	6.64e+01
20	43	76.8	667	2	F70682	probable membrane pro	6.64e+01
21	43	76.8	746	2	T06017	subtilisin homolog -	6.64e+01
22	43	76.8	773	2	T00502	protein kinase homolo	6.64e+01
23	42	75.0	110	1	VHBPDL	major capsid protein	1.04e+02
24	42	75.0	216	1	C41316	flagellin B2 precurso	1.04e+02
25	42	75.0	218	1	B41316	flagellin B1 precurso	1.04e+02
26	42	75.0	264	2	A71367	probable holocytochro	1.04e+02
27	42	75.0	303	2	D70955	hypothetical protein	1.04e+02
28	42	75.0	308	2	H64153	hypothetical protein	1.04e+02
29	42	75.0	314	2	G70046	iron-binding protein	1.04e+02
30	42	75.0	333	1	DEALXE	acetoin[2,6-dichlorop	1.04e+02
31	42	75.0	365	2	C72308	conserved hypothetica	1.04e+02
32	42	75.0	367	2	T04521	proline-rich protein	1.04e+02
33	42	75.0	406	2	H64925	aminotransferase nifS	1.04e+02
34	42	75.0	411	2	D75095	probable polysacchari	1.04e+02
35	42	75.0	417	2	D70321	sulfide dehydrogenase	1.04e+02
36	42	75.0	461	2	T03561	hypothetical protein	1.04e+02
37	42	75.0	464	2	C70414	NADH dehydrogenase (u	1.04e+02
38	42	75.0	501	2	S18573	L-lysine transport pr	1.04e+02
39	42	75.0	503	1	F64713	protein-export membra	1.04e+02
40	42	75.0	526	2	D71805	protein-export membra	1.04e+02
41	42	75.0	530	2	S61034	AIP2 protein - yeast	1.04e+02
42	42	75.0	659	2	S65976	hypothetical protein	1.04e+02
43	42	75.0	885	2	B69783	transporter homolog y	1.04e+02
44	42	75.0	1325	2	A64905	ydeK protein - Escher	1.04e+02
45	42	75.0	1530	2	S52239	hba2 protein - fissio	1.04e+02

ALIGNMENTS

```
RESULT    1
ENTRY             A55253     #type complete
TITLE             melanoma antigen MART-1 - human
ALTERNATE_NAMES   melan-A protein
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change
                    10-Sep-1997
ACCESSIONS        A55253; I38506
REFERENCE         A55253
   #authors       Kawakami, Y.; Eliyahu, S.; Delgado, C.H.; Robbins, P.F.;
                    Rivoltini, L.; Topalian, S.L.; Miki, T.; Rosenberg, S.A.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1994) 91:3515-3519
   #title         Cloning of the gene coding for a shared human melanoma
                    antigen recognized by autologous T cells infiltrating into
                    tumor.
   #cross-references MUID:94224770
   #accession     A55253
      ##status         preliminary
      ##molecule_type mRNA
      ##residues       1-118 ##label KAW
      ##cross-references GB:U06452; NID:g476131; PID:g476132
REFERENCE         I38506
   #authors       Coulie, P.G.; Brichard, V.; Van Pel, A.; Wolfel, T.;
                    Schneider, J.; Traversari, C.; Mattei, S.; De Plaen, E.;
                    Lurquin, C.; Szikora, J.P.; Renauld, J.; Boon, T.
   #journal       J. Exp. Med. (1994) 180:35-42
   #title         A new gene coding for a differentiation antigen recognized by
                    autologous cytolytic T lymphocytes on HLA-A2 melanomas [see
                    comments].
   #cross-references MUID:94275389
   #accession     I38506
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-118 ##label RES
      ##cross-references EMBL:U06654; NID:g517022; PID:g517023
GENETICS
   #gene          GDB:MLANA
      ##cross-references GDB:358979
   #map_position 17q21-17q24
SUMMARY           #length 118  #molecular-weight 13157  #checksum 3535

  Query Match          100.0%;  Score 56;  DB 2;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 9.43e-02;
  Matches     9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      27 AAGIGILTV 35
           |||||||||
```



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 21:46:46 2000; MasPar time 6.10 Seconds
577.758 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-2
Description: (1-118) from US09267439.pep
Perfect Score: 889
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 41.234; Variance 64.540; scale 0.639

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	889	100.0	118	1	MAR1_HUMAN	MELANOMA ANTIGEN RECOG	2.48e-198
2	99	11.1	344	1	CD2_MOUSE	T-CELL SURFACE ANTIGEN	4.28e-03
3	90	10.1	628	1	LU_HUMAN	LUTHERAN BLOOD GROUP G	1.28e-01
4	88	9.9	942	1	TMK1_ARATH	PUTATIVE RECEPTOR PROT	2.64e-01
5	87	9.8	519	1	TYR2_HUMAN	DOPACHROME TAUTOMERASE	3.78e-01
6	87	9.8	1630	1	PTP1_DROME	PROTEIN-TYROSINE PHOSP	3.78e-01
7	86	9.7	517	1	TYR2_MOUSE	DOPACHROME TAUTOMERASE	5.39e-01
8	86	9.7	704	1	MEPB_MOUSE	MEPRIN A BETA-SUBUNIT	5.39e-01
9	86	9.7	774	1	HN4G_HUMAN	HEPATOCYTE NUCLEAR FAC	5.39e-01
10	85	9.6	320	1	YB9K_YEAST	HYPOTHETICAL 36.0 KD P	7.66e-01
11	85	9.6	704	1	MEPB_RAT	MEPRIN A BETA-SUBUNIT	7.66e-01
12	85	9.6	918	1	PMA1_YEAST	PLASMA MEMBRANE ATPASE	7.66e-01
13	84	9.4	327	1	CD1A_HUMAN	T-CELL SURFACE GLYCOPR	1.09e+00
14	84	9.4	334	1	Y472_RICPR	HYPOTHETICAL PROTEIN R	1.09e+00
15	83	9.3	306	1	CD80_MOUSE	T LYMPHOCYTE ACTIVATIO	1.53e+00
16	83	9.3	348	1	SKI_MOUSE	SKI ONCOGENE (C-SKI) (	1.53e+00
17	83	9.3	700	1	MEPB_HUMAN	MEPRIN A BETA-SUBUNIT	1.53e+00
18	82	9.2	426	1	UCR2_SCHPO	UBIQUINOL-CYTOCHROME C	2.16e+00
19	82	9.2	477	1	INGR_MOUSE	INTERFERON-GAMMA RECEP	2.16e+00
20	82	9.2	499	1	ANSP_ECOLI	L-ASPARAGINE PERMEASE	2.16e+00
21	82	9.2	897	1	CYRB_HUMAN	CYTOKINE RECEPTOR COMM	2.16e+00
22	82	9.2	1010	1	SNI2_YEAST	SNI2 PROTEIN (SRO77 PR	2.16e+00
23	82	9.2	1013	1	PTPX_MACNE	PROTEIN-TYROSINE PHOSP	2.16e+00
24	82	9.2	1015	1	PTPX_HUMAN	PROTEIN-TYROSINE PHOSP	2.16e+00
25	82	9.2	1807	1	ITB4_RAT	INTEGRIN BETA-4 PRECUR	2.16e+00
26	81	9.1	244	1	YK91_MYCTU	HYPOTHETICAL 26.0 KD P	3.02e+00
27	81	9.1	511	1	DOP1_DROME	DOPAMINE RECEPTOR 1 PR	3.02e+00
28	81	9.1	553	1	MCP2_SALTY	METHYL-ACCEPTING CHEMO	3.02e+00
29	81	9.1	623	1	PTR2_CANAL	PEPTIDE TRANSPORTER PT	3.02e+00
30	81	9.1	855	1	ENV_FIVU2	ENV POLYPROTEIN PRECUR	3.02e+00
31	81	9.1	856	1	ENV_FIVPE	ENV POLYPROTEIN PRECUR	3.02e+00
32	81	9.1	1030	1	VPP1_CAEEL	PUTATIVE CLATHRIN-COAT	3.02e+00
33	80	9.0	307	1	YAC2_SCHPO	HYPOTHETICAL 33.9 KD P	4.23e+00
34	79	8.9	128	1	MF18_MAIZE	MFS18 PROTEIN PRECURSO	5.89e+00
35	79	8.9	135	1	RL15_HELPY	50S RIBOSOMAL PROTEIN	5.89e+00
36	79	8.9	182	1	RL11_ORYSA	60S RIBOSOMAL PROTEIN	5.89e+00
37	79	8.9	248	1	MYP0_HUMAN	MYELIN P0 PROTEIN PREC	5.89e+00
38	79	8.9	446	1	HN4B_XENLA	HEPATOCYTE NUCLEAR FAC	5.89e+00
39	79	8.9	468	1	CB1A_FUGRU	CANNABINOID RECEPTOR T	5.89e+00
40	79	8.9	546	1	VGLF_RINDL	FUSION GLYCOPROTEIN PR	5.89e+00
41	79	8.9	552	1	DPIB_ECOLI	SENSOR KINASE DPIB (EC	5.89e+00
42	79	8.9	846	1	ITBX_DROME	INTEGRIN BETA-SUBUNIT	5.89e+00
43	79	8.9	854	1	ENV_FIVSD	ENV POLYPROTEIN PRECUR	5.89e+00
44	79	8.9	1184	1	BIMC_EMENI	KINESIN-LIKE PROTEIN B	5.89e+00
45	78	8.8	98	1	ELIB_PHYDR	BETA-ELICITIN DRE-BETA	8.17e+00

ALIGNMENTS

```
RESULT    1
ID   MAR1_HUMAN     STANDARD;      PRT;   118 AA.
AC   Q16655;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MELANOMA ANTIGEN RECOGNIZED BY T-CELLS 1 (MART-1) (MELAN-A PROTEIN)
DE   (ANTIGEN SK29-AA) (ANTIGEN LB39-AA).
GN   MLANA OR MART1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RX   MEDLINE; 94224770.
RA   KAWAKAMI Y., ELIYAHU S., DELGADO C.H., ROBBINS P.F., RIVOLTINI L.,
RA   TOPALIAN S.L., MIKI T., ROSENBERG S.A.;
RT   "Cloning of the gene coding for a shared human melanoma antigen
RT   recognized by autologous T cells infiltrating into tumor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3515-3519(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94275389.
RA   COULIE P.G., BRICHARD V., VAN PEL A., WOELFEL T., SCHNEIDER J.,
RA   TRAVERSARI C., MATTEI S., DE PLAEN E., LURQUIN C., SZIKORA J.-P.,
RA   RENAULD J.-C., BOON T.;
RT   "A new gene coding for a differentiation antigen recognized by
RT   autologous cytolytic T lymphocytes on HLA-A2 melanomas.";
RL   J. Exp. Med. 180:35-42(1994).
CC   -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO MELANOMA AND
CC       MELANOCYTE CELL LINES AND RETINA.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U06452; AAA19238.1; -.
DR   EMBL; U06654; AAA20389.1; -.
KW   Antigen; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
SQ   SEQUENCE   118 AA;  13157 MW;  DFE2CFC6 CRC32;
```

```
  Query Match          100.0%;  Score 889;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 2.48e-198;
  Matches  118;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Db      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Db     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT    2
ID    CD2_MOUSE      STANDARD;      PRT;   344 AA.
AC    P08920;
DT    01-NOV-1988 (Rel. 09, Created)
DT    01-NOV-1988 (Rel. 09, Last sequence update)
DT    01-FEB-1996 (Rel. 33, Last annotation update)
DE    T-CELL SURFACE ANTIGEN CD2 PRECURSOR (T-CELL SURFACE ANTIGEN
DE    T11/LEU-5) (LFA-2) (LFA-3 RECEPTOR).
GN    CD2.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC    Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=B10.A;
RX    MEDLINE; 87276135.
RA    SEWELL W.A., BROWN M.H., FINK P.J., KOZAK C.A., CRUMPTON M.J.;
RT    "The murine homologue of the T lymphocyte CD2 antigen: molecular
RT    cloning, chromosome assignment and cell surface expression.";
RL    Eur. J. Immunol. 17:1015-1020(1987).
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE; 88004738.
RA    CLAYTON L.K., SAYRE P.H., NOVOTNY J., REINHERZ E.L.;
RT    "Murine and human T11 (CD2) cDNA sequences suggest a common signal
RT    transduction mechanism.";
RL    Eur. J. Immunol. 17:1367-1370(1987).
RN    [3]
RP    SEQUENCE FROM N.A.
RC    STRAIN=BALB/C; TISSUE=LIVER;
RX    MEDLINE; 88144486.
RA    DIAMOND D.J., CLAYTON L.K., SAYRE P.H., REINHERZ E.L.;
RT    "Exon-intron organization and sequence comparison of human and murine
RT    T11 (CD2) genes.";
RL    Proc. Natl. Acad. Sci. U.S.A. 85:1615-1619(1988).
CC    -!- FUNCTION: CD2 INTERACTS WITH LYMPHOCYTE FUNCTION-ASSOCIATED
CC        ANTIGEN (LFA-3) AND OX-45/BCM-1 TO MEDIATE ADHESION BETWEEN T
CC        CELLS AND OTHER CELL TYPES. CD2 IS IMPLICATED IN THE TRIGGERRING
CC        OF T-CELLS, THE CYTOPLASMIC DOMAIN IS IMPLICATED IN THE
CC        SIGNALING FUNCTION.
CC    -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC    -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC    ----------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ----------------------------------------------------------------------
DR    EMBL; Y00023; CAA68258.1; -.
DR    EMBL; X06143; CAA29500.1; -.
DR    EMBL; M19807; AAA37393.1; -.
DR    EMBL; M19799; AAA37393.1; JOINED.
DR    EMBL; M19801; AAA37393.1; JOINED.
DR    EMBL; M19803; AAA37393.1; JOINED.
DR    EMBL; M19805; AAA37393.1; JOINED.
DR    PIR; S02293; S02293.
DR    PIR; S01347; S01347.
DR    PIR; B28967; B28967.
DR    HSSP; P08921; 1A7B.
DR    MGD; MGI:88320; CD2.
KW    Immunoglobulin domain; T-cell; Glycoprotein; Antigen; Transmembrane;
KW    Cell adhesion; Signal.
FT    SIGNAL        1     22
FT    CHAIN        23    344       T-CELL SURFACE ANTIGEN CD2.
FT    DOMAIN       23    202       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM    204    229       POTENTIAL.
FT    DOMAIN      230    344       CYTOPLASMIC (POTENTIAL).
FT    DOMAIN       23    121       IG-LIKE V-TYPE DOMAIN.
FT    DOMAIN      122    202       IG-LIKE C2-TYPE DOMAIN.
FT    DOMAIN      276    343       PRO-RICH.
FT    DISULFID    133    197       BY SIMILARITY.
FT    DISULFID    140    180       BY SIMILARITY.
FT    CARBOHYD     82     82       POTENTIAL.
FT    CARBOHYD     94     94       POTENTIAL.
FT    CARBOHYD    135    135       POTENTIAL.
FT    CARBOHYD    166    166       POTENTIAL.
FT    CONFLICT     99     99       Y -> T (IN REF. 3).
FT    CONFLICT    128    128       M -> V (IN REF. 3).
FT    CONFLICT    175    175       N -> A (IN REF. 3).
FT    CONFLICT    191    191       K -> N (IN REF. 2).
FT    CONFLICT    192    192       M -> T (IN REF. 3).
SQ    SEQUENCE   344 AA;  38414 MW;  3064BF86 CRC32;

  Query Match          11.1%;  Score 99;  DB 1;  Length 344;
  Best Local Similarity 40.4%;  Pred. No. 4.28e-03;
  Matches   21;  Conservative    9;  Mismatches  18;  Indels   4;  Gaps   4;

Db    198 PEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEELEI 245
          | || :  | |   :|| |:| |:| | |:| |  |:||:  |  :|  | :
Qy     13 PKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSLHV 64


RESULT    3
ID    LU_HUMAN       STANDARD;      PRT;   628 AA.
AC    P50895;
DT    01-OCT-1996 (Rel. 34, Created)
DT    01-OCT-1996 (Rel. 34, Last sequence update)
DT    01-NOV-1997 (Rel. 35, Last annotation update)
DE    LUTHERAN BLOOD GROUP GLYCOPROTEIN PRECURSOR (B-CAM CELL SURFACE
DE    GLYCOPROTEIN) (AUBERGER B ANTIGEN) (F8/G253 ANTIGEN).
GN    LU OR BCAM OR MSK19.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC    Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN    [1]
RP    SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.
RC    TISSUE=PLACENTA;
RX    MEDLINE; 95296337.
RA    PARSONS S.F., MALLINSON G., HOLMES C.H., HOULIHAN J.M., SIMPSON K.L.,
RA    MAWBY W.J., SPURR N.K., WARNE D., BARCLAY A.N., ANSTEE D.J.;
RT    "The Lutheran blood group glycoprotein, another member of the
RT    immunoglobulin superfamily, is widely expressed in human tissues and
RT    is developmentally regulated in human liver.";
RL    Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).
RN    [2]
RP    SEQUENCE OF 1-588 FROM N.A.
RX    MEDLINE; 95042297.
RA    CAMPBELL I.G., FOULKES W.D., SENGER G., TROWSDALE J.,
RA    GARIN-CHESA P., RETTIG W.J.;
RT    "Molecular cloning of the B-CAM cell surface glycoprotein of
RT    epithelial cancers: a novel member of the immunoglobulin
RT    superfamily.";
RL    Cancer Res. 54:5761-5765(1994).
CC    -!- FUNCTION: PROBABLE RECEPTOR. MAY MEDIATE INTRAELLULAR SIGNALING.
CC    -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC    -!- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION (HIGHEST IN THE
CC        PANCREAS AND VERY LOW IN BRAIN). CLOSELY ASSOCIATED WITH THE BASAL
CC        LAYER OF CELLS IN EPITHELIA AND THE ENDOTHELIUM OF BLOOD VESSEL
CC        WALLS.
CC    -!- DEVELOPMENTAL STAGE: IS UNDER DEVELOPMENTAL CONTROL IN LIVER AND
```

```
CC       MAY ALSO BE REGULATED DURING DIFFERENTIATION IN OTHER TISSUES.
CC       UPREGULATED FOLLOWING MALIGNANT TRANSFORMATION IN SOME CELL TYPES.
CC   -!- POLYMORPHISM: LU IS RESPONSIBLE FOR THE LUTHERAN BLOOD GROUP
CC       SYSTEM.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       3 C2-LIKE AND 2 V-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83425; CAA58449.1; -.
DR   EMBL; X80026; CAA56327.1; -.
DR   MIM; 111200; -.
DR   PFAM; PF00047; ig; 4.
KW   Receptor; Immunoglobulin domain; Glycoprotein; Transmembrane; Signal;
KW   Blood group antigen.
FT   SIGNAL       1     31
FT   CHAIN       32    628       LUTHERAN BLOOD GROUP GLYCOPROTEIN.
FT   DOMAIN      32    547       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   548    568       POTENTIAL.
FT   DOMAIN     569    628       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      46    132       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     165    244       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN     284    344       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     377    431       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     466    529       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    53    125       PROBABLE.
FT   DISULFID   172    237       PROBABLE.
FT   DISULFID   291    337       PROBABLE.
FT   DISULFID   384    424       PROBABLE.
FT   DISULFID   473    522       PROBABLE.
FT   CARBOHYD   321    321       POTENTIAL.
FT   CARBOHYD   377    377       POTENTIAL.
FT   CARBOHYD   383    383       POTENTIAL.
FT   CARBOHYD   419    419       POTENTIAL.
FT   CARBOHYD   439    439       POTENTIAL.
FT   CONFLICT   225    226       RL -> PC (IN REF. 2).
FT   CONFLICT   355    356       EL -> DV (IN REF. 2).
SQ   SEQUENCE   628 AA;  67374 MW;  2D577E45 CRC32;

  Query Match           10.1%;  Score 90;  DB 1;  Length 628;
  Best Local Similarity  34.4%;  Pred. No. 1.28e-01;
  Matches   11;  Conservative   12;  Mismatches   8;  Indels   1;  Gaps   1;

Db    544 TSQAGVAVMAVAVSVGLLLLVVAVFYCVRRKG 575
          |:: : ::  ::| :|:|||: :  || ||:|
Qy     23 TAEEAAGIGILTVILGVLLLI-GCWYCRRRNG 53


RESULT    4
ID   TMK1_ARATH     STANDARD;      PRT;   942 AA.
AC   P43298;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PUTATIVE RECEPTOR PROTEIN KINASE TMK1 PRECURSOR (EC 2.7.1.-).
GN   TMK1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE; 93076110.
RA   CHANG C., SCHALLER G.E., PATTERSON S.E., KWOK S.F.,
RA   MEYEROWITZ E.M., BLEECKER A.B.;
RT   "The TMK1 gene from Arabidopsis codes for a protein with structural
RT   and biochemical characteristics of a receptor protein kinase.";
RL   Plant Cell 4:1263-1271(1992).
CC   -!- FUNCTION: PROBABLE RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC   -!- PTM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.
CC   -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC       PROTEIN KINASES.
CC   -!- SIMILARITY: THE REPEATED LEUCINE-RICH (LRR) SEGMENT IS FOUND IN
CC       MANY PROTEINS. NUMBER IN THIS PROTEIN: 11.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00670; AAA32876.1; -.
DR   HSSP; P11362; 1FGI.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   PFAM; PF00560; LRR; 8.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Transmembrane; Receptor; Glycoprotein; Signal; Repeat;
KW   Leucine-repeat; Phosphorylation.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24    942       PUTATIVE RECEPTOR PROTEIN KINASE TMK1.
FT   DOMAIN      24    482       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   483    503       POTENTIAL.
FT   DOMAIN     504    942       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     588    869       PROTEIN KINASE.
FT   NP_BIND    594    602       ATP (BY SIMILARITY).
FT   BINDING    616    616       ATP (BY SIMILARITY).
FT   ACT_SITE   717    717       BY SIMILARITY.
FT   CARBOHYD    86     86       POTENTIAL.
FT   CARBOHYD    99     99       POTENTIAL.
FT   CARBOHYD   158    158       POTENTIAL.
FT   CARBOHYD   164    164       POTENTIAL.
FT   CARBOHYD   171    171       POTENTIAL.
FT   CARBOHYD   230    230       POTENTIAL.
FT   CARBOHYD   285    285       POTENTIAL.
FT   CARBOHYD   363    363       POTENTIAL.
FT   CARBOHYD   397    397       POTENTIAL.
SQ   SEQUENCE   942 AA;  102387 MW;  79E1A873 CRC32;

  Query Match            9.9%;  Score 88;  DB 1;  Length 942;
  Best Local Similarity  37.5%;  Pred. No. 2.64e-01;
  Matches    9;  Conservative   10;  Mismatches   4;  Indels   1;  Gaps   1;

Db    491 GLLSIFL-IGLLVFCWYKKRQKRF 513
          |:|:::| : ||: ||| :|:: :
Qy     31 GILTVILGVLLLIGCWYCRRRNGY 54


RESULT    5
ID   TYR2_HUMAN     STANDARD;      PRT;   519 AA.
AC   P40126;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   DOPACHROME TAUTOMERASE PRECURSOR (EC 5.3.3.12) (DT) (DCT) (DOPACHROME
DE   DELTA-ISOMERASE) (TYROSINASE-RELATED PROTEIN 2) (TRP-2) (TRP2).
GN   DCT OR TYRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE; 94198295.
RA   YOKOYAMA K., SUZUKI H., YASUMOTO K.I., TOMITA Y., SHIBAHARA S.;
RT   "Molecular cloning and functional analysis of a cDNA coding for human
RT   DOPAchrome tautomerase/tyrosinase-related protein-2.";
RL   Biochim. Biophys. Acta 1217:317-321(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94266170.
RA   CASSADY J.L., STURM R.A.;
RT   "Sequence of the human dopachrome tautomerase-encoding TRP-2 cDNA.";
RL   Gene 143:295-298(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94139684.
RA   BOUCHARD B., DEL MARMOL V., JACKSON I.J., CHERIF D., DUBERTRET L.;
RT   "Molecular characterization of a human tyrosinase-related-protein-2
RT   cDNA. Patterns of expression in melanocytic cells.";
RL   Eur. J. Biochem. 219:127-134(1994).
RN   [4]
RP   SEQUENCE OF 1-98 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 96079088.
RA   STURM R.A., O'SULLIVAN B.J., BOX N.F., SMITH A.G., SMIT S.E.,
RA   PUTTICK E.R., PARSONS P.G., DUNN I.S.;
RT   "Chromosomal structure of the human TYRP1 and TYRP2 loci and
RT   comparison of the tyrosinase-related protein gene family.";
RL   Genomics 29:24-34(1995).
RN   [5]
RP   SEQUENCE OF 1-98 FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95014579.
RA   YOKOYAMA K., YASUMOTO K.I., SUZUKI H., SHIBAHARA S.;
RT   "Cloning of the human DOPAchrome tautomerase/tyrosinase-related
RT   protein 2 gene and identification of two regulatory regions required
RT   for its pigment cell-specific expression.";
RL   J. Biol. Chem. 269:27080-27087(1994).
CC   -!- CATALYTIC ACTIVITY: L-DOPACHROME = 5,6-DIHYDROXYINDOLE-2-
CC       CARBOXYLATE.
CC   -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC   -!- PATHWAY: MELANIN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MELANOSOMAL.
CC   -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D17547; BAA04484.1; -.
DR   EMBL; L18967; AAA20870.1; -.
DR   EMBL; S69231; AAC60627.1; -.
DR   EMBL; L38953; AAC41925.1; -.
DR   EMBL; D28767; BAA05956.1; -.
DR   PIR; S43510; S43510.
DR   MIM; 191275; -.
DR   PROSITE; PS00497; TYROSINASE_1; 1.
DR   PROSITE; PS00498; TYROSINASE_2; 1.
DR   PFAM; PF00264; tyrosinase; 1.
KW   Isomerase; Zinc; Glycoprotein; Signal; Transmembrane;
KW   Melanin biosynthesis.
FT   SIGNAL       1     23       POTENTIAL.
FT   CHAIN       24    519       DOPACHROME TAUTOMERASE.
FT   DOMAIN      24    472       LUMENAL, MELANOSOME (POTENTIAL).
FT   TRANSMEM   473    493       POTENTIAL.
FT   DOMAIN     494    519       CYTOPLASMIC (POTENTIAL).
FT   METAL      189    189       ZINC A (BY SIMILARITY).
FT   METAL      211    211       ZINC A (BY SIMILARITY).
FT   METAL      220    220       ZINC A (BY SIMILARITY).
FT   METAL      369    369       ZINC B (BY SIMILARITY).
FT   METAL      373    373       ZINC B (BY SIMILARITY).
FT   METAL      396    396       ZINC B (BY SIMILARITY).
FT   CARBOHYD   170    170       POTENTIAL.
FT   CARBOHYD   178    178       POTENTIAL.
FT   CARBOHYD   237    237       POTENTIAL.
FT   CARBOHYD   300    300       POTENTIAL.
FT   CARBOHYD   342    342       POTENTIAL.
FT   CARBOHYD   377    377       POTENTIAL.
SQ   SEQUENCE   519 AA;  59145 MW;  4FEFCCD2 CRC32;

  Query Match            9.8%;  Score 87;  DB 1;  Length 519;
  Best Local Similarity  35.3%;  Pred. No. 3.78e-01;
  Matches    12;  Conservative   11;  Mismatches  10;  Indels   1;  Gaps   1;

Db    477 MGTLVALVGLFVLLAFLQYRRLRKGYTPLMETHL 510
          :| |  ::|:::|::    || |:|| :||:  |
Qy     30 IGILTVILGVLLLIGCWYCRR-RNGYRALMDKSL 62


RESULT    6
ID   PTP1_DROME     STANDARD;      PRT;  1630 AA.
AC   P35992;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   PROTEIN-TYROSINE PHOSPHATASE 10D PRECURSOR (EC 3.1.3.48) (RECEPTOR-
DE   LINKED PROTEIN-TYROSINE PHOSPHATASE 10D).
GN   PTP10D.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 92034989.
RA   TIAN S.-S., TSOULFAS P., ZINN K.;
RT   "Three receptor-linked protein-tyrosine phosphatases are selectively
RT   expressed on central nervous system axons in the Drosophila embryo.";
RL   Cell 67:675-685(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE; 92034988.
RA   YANG X., SEOW K.T., BAHRI S.M., OON S.H., CHIA W.;
RT   "Two Drosophila receptor-like tyrosine phosphatase genes are
RT   expressed in a subset of developing axons and pioneer neurons in the
RT   embryonic CNS.";
RL   Cell 67:661-673(1991).
CC   -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC       PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: TWO ISOFORMS THAT DIFFER IN THEIR C-TERMINAL
CC       TAILS ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND
CC       PIONEER NEURONS IN THE EMBRYO.
CC   -!- SIMILARITY: 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 12 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M80538; AAA28952.1; -.
DR   EMBL; M80465; AAA28484.1; -.
DR   PIR; C41214; C41214.
DR   PIR; D41214; D41214.
DR   PIR; A41215; A41215.
DR   HSSP; P18052; 1YFO.
DR   FLYBASE; FBgn0004370; Ptp10D.
```

```
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
DR   PFAM; PF00041; fn3; 10.
DR   PFAM; PF00102; Y_phosphatase; 1.
KW   Transmembrane; Hydrolase; Duplication; Signal; Alternative splicing;
KW   Repeat.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35   1630       PROTEIN-TYROSINE PHOSPHATASE 10D.
FT   DOMAIN       35   1196       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1197   1218       POTENTIAL.
FT   DOMAIN     1219   1630       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       43    119       FIBRONECTIN TYPE-III.
FT   DOMAIN      120    214       FIBRONECTIN TYPE-III.
FT   DOMAIN      215    308       FIBRONECTIN TYPE-III.
FT   DOMAIN      309    402       FIBRONECTIN TYPE-III.
FT   DOMAIN      403    493       FIBRONECTIN TYPE-III.
FT   DOMAIN      494    580       FIBRONECTIN TYPE-III.
FT   DOMAIN      581    669       FIBRONECTIN TYPE-III.
FT   DOMAIN      670    766       FIBRONECTIN TYPE-III.
FT   DOMAIN      767    861       FIBRONECTIN TYPE-III.
FT   DOMAIN      862    955       FIBRONECTIN TYPE-III.
FT   DOMAIN      956   1048       FIBRONECTIN TYPE-III.
FT   DOMAIN     1049   1189       FIBRONECTIN TYPE-III.
FT   DOMAIN     1290   1532       PROTEIN-TYROSINE PHOSPHATASE.
FT   CARBOHYD     75     75       POTENTIAL.
FT   CARBOHYD    106    106       POTENTIAL.
FT   CARBOHYD    128    128       POTENTIAL.
FT   CARBOHYD    169    169       POTENTIAL.
FT   CARBOHYD    212    212       POTENTIAL.
FT   CARBOHYD    229    229       POTENTIAL.
FT   CARBOHYD    259    259       POTENTIAL.
FT   CARBOHYD    289    289       POTENTIAL.
FT   CARBOHYD    317    317       POTENTIAL.
FT   CARBOHYD    471    471       POTENTIAL.
FT   CARBOHYD    486    486       POTENTIAL.
FT   CARBOHYD    512    512       POTENTIAL.
FT   CARBOHYD    533    533       POTENTIAL.
FT   CARBOHYD    588    588       POTENTIAL.
FT   CARBOHYD    668    668       POTENTIAL.
FT   CARBOHYD    687    687       POTENTIAL.
FT   CARBOHYD    719    719       POTENTIAL.
FT   CARBOHYD    723    723       POTENTIAL.
FT   CARBOHYD    823    823       POTENTIAL.
FT   CARBOHYD    841    841       POTENTIAL.
FT   CARBOHYD    874    874       POTENTIAL.
FT   CARBOHYD    908    908       POTENTIAL.
FT   CARBOHYD    925    925       POTENTIAL.
FT   CARBOHYD   1001   1001       POTENTIAL.
FT   CARBOHYD   1104   1104       POTENTIAL.
FT   CARBOHYD   1135   1135       POTENTIAL.
FT   CARBOHYD   1194   1194       POTENTIAL.
FT   ACT_SITE   1467   1467       BY SIMILARITY.
FT   VARSPLIC   1548   1557       GQQVQLDENG -> DDEGIAESGM (IN SHORT
FT                                ISOFORM).
FT   VARSPLIC   1558   1630       MISSING (IN SHORT ISOFORM).
FT   CONFLICT    124    124       D -> I (IN REF. 2).
FT   CONFLICT    127    127       S -> L (IN REF. 2).
FT   CONFLICT   1125   1125       Y -> YQ (IN REF. 1).
FT   CONFLICT   1166   1167       IG -> YR (IN REF. 1).
FT   CONFLICT   1172   1172       R -> A (IN REF. 1).
FT   CONFLICT   1216   1216       I -> L (IN REF. 1).
FT   CONFLICT   1457   1457       C -> G (IN REF. 1).
SQ   SEQUENCE   1630 AA;  184861 MW;  06B7B9FE CRC32;

  Query Match           9.8%;  Score 87;  DB 1;  Length 1630;
  Best Local Similarity  37.0%;  Pred. No. 3.78e-01;
  Matches   17;  Conservative    7;  Mismatches  19;  Indels   3;  Gaps   3;

Db   1184 YSFPIQTD-QDNTSLIVAITVP-LTIIL-VLLVTLIFYKRRRNNCR 1226
           |::|     :  |:   |  :  ||:|| |||:    | ||||  |
Qy     10 YGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYR 55
```

```
RESULT    7
ID   TYR2_MOUSE     STANDARD;      PRT;   517 AA.
AC   P29812;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   DOPACHROME TAUTOMERASE PRECURSOR (EC 5.3.3.12) (DT) (DCT) (DOPACHROME
DE   DELTA-ISOMERASE) (TYROSINASE-RELATED PROTEIN 2) (TRP-2) (TRP2) (SLATY
DE   LOCUS PROTEIN).
GN   DCT OR TYRP2 OR TYRP-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 92164640.
RA   JACKSON I.J., CHAMBERS D.M., TSUKAMOTO K., COPELAND N.G.,
RA   GILBERT D.J., JENKINS N.A., HEARING V.J.;
RT   "A second tyrosinase-related protein, TRP-2, maps to and is mutated
RT   at the mouse slaty locus.";
RL   EMBO J. 11:527-536(1992).
RN   [2]
RP   SEQUENCE OF 1-98 FROM N.A., AND VARIANTS SLATY-2J AND SLATY-LT.
RC   STRAIN=129/SV;
RX   MEDLINE; 96079089.
RA   BUDD P.S., JACKSON I.J.;
RT   "Structure of the mouse tyrosinase-related protein-2/dopachrome
RT   tautomerase (Tyrp2/Dct) gene and sequence of two novel slaty
RT   alleles.";
RL   Genomics 29:35-43(1995).
RN   [3]
RP   CHARACTERIZATION.
RX   MEDLINE; 92164639.
RA   TSUKAMOTO K., JACKSON I.J., URABE K., MONTAGUE P.M., HEARING V.J.;
RT   "A second tyrosinase-related protein, TRP-2, is a melanogenic enzyme
RT   termed DOPAchrome tautomerase.";
RL   EMBO J. 11:519-526(1992).
RN   [4]
RP   ZINC-BINDING.
RX   MEDLINE; 95071460.
RA   SOLANO F., MARTINEZ-LIARTE J.H., JIMENEZ-CERVANTES C.,
RA   GARCIA-BORRON J.C., LOZANO J.A.;
RT   "Dopachrome tautomerase is a zinc-containing enzyme.";
RL   Biochem. Biophys. Res. Commun. 204:1243-1250(1994).
RN   [5]
RP   ZINC-BINDING.
RX   MEDLINE; 96152526.
RA   SOLANO F., JIMENEZ-CERVANTES C., MARTINEZ-LIARTE J.H.,
RA   GARCIA-BORRON J.C., JARA J.R., LOZANO J.A.;
RT   "Molecular mechanism for catalysis by a new zinc-enzyme, dopachrome
RT   tautomerase.";
RL   Biochem. J. 313:447-453(1996).
CC   -!- CATALYTIC ACTIVITY: L-DOPACHROME = 5,6-DIHYDROXYINDOLE-2-
CC       CARBOXYLATE.
CC   -!- COFACTOR: BINDS TWO ZINC IONS.
CC   -!- PATHWAY: MELANIN BIOSYNTHESIS.
CC   -!- SUBUNIT: TYROSINASE, TRP-1 AND TRP-2 MAY FORM A MULTIENZYME
CC       COMPLEX.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MELANOSOMAL.
CC   -!- TISSUE SPECIFICITY: MELANOCYTES AND RETINAL PIGMENTED EPITHELIUM.
CC   -!- DISEASE: THE SLATY MUTATION IN TYRP2 LEADS TO A DECREASE OF DT
CC       ACTIVITY AND A CONSEQUENT CHANGE IN THE PIGMENTATION OF THE MICE
CC       TO A DARK GREY/BROWN EUMELANIN. THE SLATY-2J MUTATION HAS A
CC       SIMILAR PHENOTYPE, THE SLATY-LT (LIGHT) MUTATION HAS A MORE SEVERE
CC       EFFECT AND IS SEMIDOMINANT, ITS PHENOTYPE MAY BE A RESULT OF THE
CC       FAILURE OF THE ENZYME TO BE CORRECTLY TARGETED TO ITS NORMAL
CC       LOCATION ON THE INNER FACE OF THE MELANOSOMAL MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63349; CAA44951.1; -.
DR   EMBL; X85126; CAA59440.1; -.
DR   PIR; S19243; S19243.
DR   MGD; MGI:102563; DCT.
DR   PROSITE; PS00497; TYROSINASE_1; 1.
DR   PROSITE; PS00498; TYROSINASE_2; 1.
DR   PFAM; PF00264; tyrosinase; 1.
KW   Isomerase; Zinc; Glycoprotein; Signal; Transmembrane;
KW   Melanin biosynthesis; Disease mutation.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24    517       DOPACHROME TAUTOMERASE.
FT   DOMAIN       24    472       LUMENAL, MELANOSOME (POTENTIAL).
FT   TRANSMEM    473    491       POTENTIAL.
FT   DOMAIN      492    517       CYTOPLASMIC (POTENTIAL).
FT   METAL       189    189       ZINC A (BY SIMILARITY).
FT   METAL       211    211       ZINC A (BY SIMILARITY).
FT   METAL       220    220       ZINC A (BY SIMILARITY).
FT   METAL       369    369       ZINC B (BY SIMILARITY).
FT   METAL       373    373       ZINC B (BY SIMILARITY).
FT   METAL       396    396       ZINC B (BY SIMILARITY).
FT   CARBOHYD     92     92       POTENTIAL.
FT   CARBOHYD    170    170       POTENTIAL.
FT   CARBOHYD    178    178       POTENTIAL.
FT   CARBOHYD    237    237       POTENTIAL.
FT   CARBOHYD    300    300       POTENTIAL.
FT   CARBOHYD    342    342       POTENTIAL.
FT   CARBOHYD    377    377       POTENTIAL.
FT   VARIANT     194    194       R -> Q (IN SLATY).
FT   VARIANT     434    434       P -> L (IN SLATY-2J).
FT   VARIANT     486    486       G -> R (IN SLATY-LT).
SQ   SEQUENCE   517 AA;  58569 MW;  8EBA0B41 CRC32;

  Query Match             9.7%;  Score 86;  DB 1;  Length 517;
  Best Local Similarity  41.2%;  Pred. No. 5.39e-01;
  Matches    14;  Conservative    9;  Mismatches  10;  Indels   1;  Gaps   1;

Db    475 IGILGAFVLLLGLLAFLQYRRLRKGYAPLMETGL 508
          ||||  :: :| |::    || |:|| :||: :|
Qy     30 IGILTVILGVLLLIGCWYCRR-RNGYRALMDKSL 62


RESULT    8
ID   MEPB_MOUSE     STANDARD;      PRT;   704 AA.
AC   Q61847;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MEPRIN A BETA-SUBUNIT PRECURSOR (EC 3.4.24.18) (ENDOPEPTIDASE-2).
GN   MEP1B OR MEP-1B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 94012651.
RA   GORBEA C.M., MARCHAND P., JIANG W., COPELAND N.G., GILBERT D.J.,
RA   JENKINS N.A., BOND J.S.;
RT   "Cloning, expression, and chromosomal localization of the mouse
RT   meprin beta subunit.";
RL   J. Biol. Chem. 268:21035-21043(1993).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM BETA').
RC   TISSUE=KIDNEY;
RX   MEDLINE; 96147211.
RA   DIETRICH J.M., BOND J.S., JIANG W.;
RT   "A novel meprin beta' mRNA in mouse embryonal and human colon
RT   carcinoma cells.";
RL   J. Biol. Chem. 271:2271-2278(1996).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEIN AND PEPTIDE SUBSTRATES
CC       PREFERENTIALLY ON CARBOXYL SIDE OF HYDROPHOBIC RESIDUES.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS WHICH
CC       IS FORMED BY THE NON-COVALENT ASSOCIATION OF TWO DISULFIDE-LINKED
CC       HETERODIMERS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS, THE BETA AND BETA' SUBUNITS,
CC       DIFFER IN THEIR N-TERMINUS DUE TO DIFFERENTIAL PROMOTER USAGE AND
CC       ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: THE BETA-SUBUNIT IS EXPRESSED IN KIDNEY,
CC       INTESTINAL BRUSH BORDERS, AND SALIVARY DUCTS. THE BETA'-ISOFORM
CC       HAS BEEN FOUND IN CARCINOMA CELLS.
CC   -!- INDUCTION: THE BETA'-SUBUNIT IS INDUCED BY THE MORPHOGEN RETINOIS
CC       ACID.
CC   -!- PTM: THIS PROTEIN UNDERGOES PROTEOLYTIC PROCESSING. BOTH FORMS
CC       ARE GLYCOSYLATED.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L15193; AAA75234.1; -.
DR   HSSP; P28825; 1IAF.
DR   MGD; MGI:96964; MEP1B.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PFAM; PF00008; EGF; 1.
DR   PFAM; PF00629; MAM; 1.
DR   PFAM; PF00917; MATH; 1.
DR   PFAM; PF01400; Astacin; 1.
KW   Transmembrane; Hydrolase; Metalloprotease; Zinc; Glycoprotein;
KW   Zymogen; Signal; EGF-like domain; Alternative splicing.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21     64       BY SIMILARITY.
FT   CHAIN        65    704       MEPRIN A BETA-SUBUNIT.
FT   DOMAIN       21    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    678       POTENTIAL.
FT   DOMAIN      679    704       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       63    260       METALLOPROTEASE.
FT   DOMAIN      261    430       MAM.
FT   DOMAIN      607    647       EGF-LIKE.
FT   METAL       153    153       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    154    154       BY SIMILARITY.
FT   METAL       157    157       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       163    163       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    611    622       BY SIMILARITY.
FT   DISULFID    616    631       BY SIMILARITY.
FT   DISULFID    633    646       BY SIMILARITY.
FT   CARBOHYD    193    193       POTENTIAL.
FT   CARBOHYD    219    219       POTENTIAL.
FT   CARBOHYD    255    255       POTENTIAL.
FT   CARBOHYD    316    316       POTENTIAL.
FT   CARBOHYD    422    422       POTENTIAL.
FT   CARBOHYD    437    437       POTENTIAL.
FT   CARBOHYD    529    529       POTENTIAL.
FT   CARBOHYD    548    548       POTENTIAL.
FT   CARBOHYD    593    593       POTENTIAL.
FT   VARSPLIC      1     27       MDARHQPWFLVFATFLLASGLPAPEKF -> MNSTAGPASR
FT                                SRHSFKCRMKLLKAPRDGMYMMTFG (IN ISOFORM
FT                                BETA').
```

```
SQ   SEQUENCE   704 AA;  79548 MW;  83CF75C1 CRC32;

  Query Match             9.7%;  Score 86;  DB 1;  Length 704;
  Best Local Similarity  26.2%;  Pred. No. 5.39e-01;
  Matches    11;  Conservative   14;  Mismatches  14;  Indels   3;  Gaps   3;

Db    648 KRGSTRDTVIIAVSSTVTVFAVML-IITLVSV-YCTRRK-YR 686
          |:|  :       :: : :::|:| :: |::  || ||: ||
Qy     14 KKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYR 55


RESULT    9
ID   HN4G_HUMAN     STANDARD;      PRT;   774 AA.
AC   Q14541;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HEPATOCYTE NUCLEAR FACTOR 4-GAMMA (HNF4-GAMMA).
GN   HNF4G OR NR2A2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 96182096.
RA   DREWES T., SENKEL S., HOLEWA B., RYFFEL G.U.;
RT   "Human hepatocyte nuclear factor 4 isoforms are encoded by distinct
RT   and differentially expressed genes.";
RL   Mol. Cell. Biol. 16:925-931(1996).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC       NR2 SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z49826; CAA89990.1; -.
DR   TRANSFAC; T02430; -.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR   PFAM; PF00104; hormone_rec; 1.
DR   PFAM; PF00105; zf-C4; 1.
KW   Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW   Zinc-finger.
FT   DNA_BIND    378    443       C4-TYPE ZINC FINGERS (TWO).
FT   ZN_FING     378    398       C4-TYPE.
FT   ZN_FING     414    438       C4-TYPE.
SQ   SEQUENCE   774 AA;  86292 MW;  A844C604 CRC32;

  Query Match             9.7%;  Score 86;  DB 1;  Length 774;
  Best Local Similarity  37.1%;  Pred. No. 5.39e-01;
  Matches    13;  Conservative    8;  Mismatches  13;  Indels   1;  Gaps   1;

Db    430 CRYCRLRKCFRAGMKKEAVQNERDRISTRRSTFDG 464
          |:||| |: :|| | | :::   :   |||   :|
Qy     45 CWYCRRRNGYRALMDK-SLHVGTQCALTRRCPQEG 78


RESULT   10
ID   YB9K_YEAST     STANDARD;      PRT;   320 AA.
AC   P38342;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   HYPOTHETICAL 36.0 KD PROTEIN IN SHM1-MRPL37 INTERGENIC REGION.
GN   YBR265W OR YBR1734.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 93220397.
RA   DOIGNON F., BITEAU N., CROUZET M., AIGLE M.;
RT   "The complete sequence of a 19,482 bp segment located on the right
RT   arm of chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 9:189-199(1993).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X70529; CAA49930.1; -.
DR   EMBL; Z36134; CAA85228.1; -.
DR   PIR; S32966; S32966.
DR   PFAM; PF00106; adh_short; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM    162    182       POTENTIAL.
FT   TRANSMEM    255    275       POTENTIAL.
FT   TRANSMEM    280    300       POTENTIAL.
SQ   SEQUENCE   320 AA;  35986 MW;  2D5C4939 CRC32;

  Query Match             9.6%;  Score 85;  DB 1;  Length 320;
  Best Local Similarity  41.7%;  Pred. No. 7.66e-01;
  Matches    10;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    287 FGVLSNILVVPFYMVGCSWYIRKW 310
          :|:|: || | : ::|| || |::
Qy     30 IGILTVILGV-LLLIGC-WYCRRR 51


RESULT   11
ID   MEPB_RAT       STANDARD;      PRT;   704 AA.
AC   P28826;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   MEPRIN A BETA-SUBUNIT PRECURSOR (EC 3.4.24.18) (ENDOPEPTIDASE-2).
GN   MEP1B.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RX   MEDLINE; 92317075.
RA   JOHNSON G.D., HERSH L.B.;
RT   "Cloning a rat meprin cDNA reveals the enzyme is a heterodimer.";
RL   J. Biol. Chem. 267:13505-13512(1992).
RN   [2]
RP   ERRATUM (RETRACTION).
RX   MEDLINE; 93359474.
RA   JOHNSON G.D., HERSH L.B.;
RL   J. Biol. Chem. 268:17647-17647(1993).
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEIN AND PEPTIDE SUBSTRATES
CC       PREFERENTIALLY ON CARBOXYL SIDE OF HYDROPHOBIC RESIDUES.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SUBUNIT: HETEROTETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS WHICH
CC       IS FORMED BY THE NON-COVALENT ASSOCIATION OF TWO DISULFIDE-LINKED
CC       HETERODIMERS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: KIDNEY, INTESTINAL BRUSH BORDERS, AND
CC       SALIVARY DUCTS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12A (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE ASTACIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M88601; AAA41587.1; -.
DR   PIR; A42908; A42908.
DR   HSSP; P28825; 1IAF.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00740; MAM_1; 1.
DR   PROSITE; PS50060; MAM_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PFAM; PF00629; MAM; 1.
DR   PFAM; PF00917; MATH; 1.
DR   PFAM; PF01400; Astacin; 1.
KW   Transmembrane; Hydrolase; Metalloprotease; Zinc; Glycoprotein;
KW   Zymogen; Signal; EGF-like domain.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21     64
FT   CHAIN        65    704       MEPRIN A BETA-SUBUNIT.
FT   DOMAIN       21    654       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    655    678       POTENTIAL.
FT   DOMAIN      679    704       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       63    260       METALLOPROTEASE.
FT   DOMAIN      261    430       MAM.
FT   DOMAIN      607    647       EGF-LIKE.
FT   METAL       153    153       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    154    154       BY SIMILARITY.
FT   METAL       157    157       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       163    163       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    611    622       BY SIMILARITY.
FT   DISULFID    616    631       BY SIMILARITY.
FT   DISULFID    633    646       BY SIMILARITY.
FT   CARBOHYD    193    193       POTENTIAL.
FT   CARBOHYD    219    219       POTENTIAL.
FT   CARBOHYD    316    316       POTENTIAL.
FT   CARBOHYD    422    422       POTENTIAL.
FT   CARBOHYD    437    437       POTENTIAL.
FT   CARBOHYD    528    528       POTENTIAL.
FT   CARBOHYD    547    547       POTENTIAL.
FT   CARBOHYD    592    592       POTENTIAL.
SQ   SEQUENCE   704 AA;  79249 MW;  CD151E13 CRC32;

  Query Match             9.6%;  Score 85;  DB 1;  Length 704;
  Best Local Similarity  28.6%;  Pred. No. 7.66e-01;
  Matches    12;  Conservative   12;  Mismatches  15;  Indels   3;  Gaps   3;

Db    648 KRGSTKDTIVIAVSSTVTVFAVML-IITLISV-YCTRRK-YR 686
          |:|          :: : :::|:| :: ||:  || ||: ||
Qy     14 KKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYR 55


RESULT   12
ID   PMA1_YEAST     STANDARD;      PRT;   918 AA.
AC   P05030;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PLASMA MEMBRANE ATPASE 1 (EC 3.6.1.35) (PROTON PUMP).
GN   PMA1 OR YGL008C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86146844.
RA   SERRANO R., KIELLAND-BRANDT M.C., FINK G.R.;
RT   "Yeast plasma membrane ATPase is essential for growth and has
RT   homology with (Na+ + K+), K+- and Ca2+-ATPases.";
RL   Nature 319:689-693(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   HEBLING U., HOFMANN B., DELIUS H.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
CC       HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
CC       ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
CC       EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKINIZATION MAY MEDIATE
CC       GROWTH RESPONSES.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- PTM: PHOSPHORYLATED ON MULTIPLE SER AND THR RESIDUES.
CC   -!- MISCELLANEOUS: THERE ARE TWO PLASMA MEMBRANE ATPASES IN YEAST.
CC       THIS IS THE MAJOR ISOFORM.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC       (E1-E2 ATPASES).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X03534; CAA27237.1; -.
DR   EMBL; Z72530; CAA96708.1; -.
DR   PIR; A25823; PXBY1P.
DR   SGD; L0001449; PMA1.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
DR   PFAM; PF00122; E1-E2_ATPase; 1.
KW   Hydrolase; Hydrogen ion transport; Transmembrane; Phosphorylation;
KW   ATP-binding; Multigene family; Glycoprotein.
FT   DOMAIN        1    115       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    116    136       1 (POTENTIAL).
FT   DOMAIN      137    140       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    141    160       2 (POTENTIAL).
FT   DOMAIN      161    291       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    292    313       3 (POTENTIAL).
FT   DOMAIN      314    325       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    326    347       4 (POTENTIAL).
FT   DOMAIN      348    719       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    720    738       5 (POTENTIAL).
FT   DOMAIN      739    754       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    755    774       6 (POTENTIAL).
FT   DOMAIN      775    824       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    825    845       7 (POTENTIAL).
FT   DOMAIN      846    857       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    858    874       8 (POTENTIAL).
FT   DOMAIN      875    918       CYTOPLASMIC (POTENTIAL).
FT   MOD_RES     378    378       PHOSPHORYLATION (PROBABLE).
FT   BINDING     474    474       ATP (BY SIMILARITY).
FT   CARBOHYD    848    848       POTENTIAL.
FT   DOMAIN        5     17       POLY-SER.
FT   DOMAIN       31     78       ASP/GLU-RICH (ACIDIC).
FT   DOMAIN       39     44       POLY-ASP.
FT   DOMAIN      585    590       POLY-GLY.
SQ   SEQUENCE   918 AA;  99619 MW;  93C1FEA3 CRC32;

  Query Match             9.6%;  Score 85;  DB 1;  Length 918;
  Best Local Similarity  34.1%;  Pred. No. 7.66e-01;
  Matches    15;  Conservative   11;  Mismatches  14;  Indels   4;  Gaps   3;

Db    282 GQGH-FTEVLNGIGIILLVLVIATLLLVWTACFY-RTNGIVRIL 323
          |:|| :|   :: || :| :::: |||:  :| | |     | |
Qy     16 GHGHSYTTAEEAAGIGILTVILGVLLLI--GCWYCRRRNGYRAL 57


RESULT   13
ID   CD1A_HUMAN     STANDARD;      PRT;   327 AA.
AC   P06126;
```

```
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   T-CELL SURFACE GLYCOPROTEIN CD1A PRECURSOR (CD1A ANTIGEN) (T-CELL
DE   SURFACE ANTIGEN T6/LEU-6) (HTA1 THYMOCYTE ANTIGEN).
GN   CD1A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 88097453.
RA   MARTIN L.H., CALABI F., LEFEBVRE F.-A., BILSLAND C.A.G., MILSTEIN C.;
RT   "Structure and expression of the human thymocyte antigens CD1a, CD1b,
RT   and CD1c.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:9189-9193(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89341413.
RA   ARUFFO A., SEED B.;
RT   "Expression of cDNA clones encoding the thymocyte antigens CD1a, b, c
RT   demonstrates a hierarchy of exclusion in fibroblasts.";
RL   J. Immunol. 143:1723-1730(1989).
RN   [3]
RP   SEQUENCE OF 99-327 FROM N.A.
RC   TISSUE=T-CELL;
RX   MEDLINE; 87014824.
RA   CALABI F., MILSTEIN C.;
RT   "A novel family of human major histocompatibility complex-related
RT   genes not mapping to chromosome 6.";
RL   Nature 323:540-543(1986).
CC   -!- FUNCTION: NOT KNOWN.
CC   -!- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON CORTICAL THYMOCYTES, ON CERTAIN
CC       T-CELL LEUKEMIAS, AND IN VARIOUS OTHER TISSUES.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD1 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd1.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28825; AAA51931.1; -.
DR   EMBL; M22167; AAA51932.1; -.
DR   EMBL; M22080; AAA51932.1; JOINED.
DR   EMBL; M22163; AAA51932.1; JOINED.
DR   EMBL; M22164; AAA51932.1; JOINED.
DR   EMBL; M22165; AAA51932.1; JOINED.
DR   EMBL; M22166; AAA51932.1; JOINED.
DR   EMBL; X04450; CAA28049.1; -.
DR   PIR; A02242; HLHUCD.
DR   PIR; A39957; A39957.
DR   PIR; A45801; A45801.
DR   HSSP; P11609; 1CD1.
DR   MIM; 188370; -.
DR   PFAM; PF00047; ig; 1.
KW   Glycoprotein; Signal; Transmembrane; Immunoglobulin domain; T-cell;
KW   Multigene family.
FT   SIGNAL       1     16
FT   CHAIN       17    327       T-CELL SURFACE GLYCOPROTEIN CD1A.
FT   DOMAIN      17    108       EXTRACELLULAR ALPHA-1.
FT   DOMAIN     109    201       EXTRACELLULAR ALPHA-2.
FT   DOMAIN     202    294       EXTRACELLULAR ALPHA-3.
FT   TRANSMEM   300    321       POTENTIAL.
FT   DOMAIN     322    327       CYTOPLASMIC (POTENTIAL).
FT   DISULFID   119    183       BY SIMILARITY.
FT   DISULFID   223    278       BY SIMILARITY.
FT   CARBOHYD    37     37       POTENTIAL.
FT   CARBOHYD    60     60       POTENTIAL.
FT   CARBOHYD    74     74       POTENTIAL.
FT   CARBOHYD   145    145       POTENTIAL.
FT   CONFLICT    30     30       I -> T (IN REF. 1).
FT   CONFLICT    68     68       W -> C (IN REF. 1).
SQ   SEQUENCE   327 AA;  37172 MW;  613EFD65 CRC32;

  Query Match             9.4%;  Score 84;  DB 1;  Length 327;
  Best Local Similarity  48.1%;  Pred. No. 1.09e+00;
  Matches    13;  Conservative    7;  Mismatches   4;  Indels   3;  Gaps   3;

Db    299 SVGFIILAVIVPLLLLIGLALWF-RKR 324
          : |: ||:||: :|||||   |: |:|
Qy     27 AAGIGILTVILGVLLLIG-C-WYCRRR 51


RESULT   14
ID   Y472_RICPR     STANDARD;      PRT;   334 AA.
AC   Q9ZD72;
DT   15-DEC-1999 (Rel. 39, Created)
DT   15-DEC-1999 (Rel. 39, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN RP472.
GN   RP472.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE; 99039499.
RA   ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,
RA   SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,
RA   ERIKSSON A.-S., WINKLER H.H., KURLAND C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235271; CAA14927.1; -.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM     46     66       POTENTIAL.
SQ   SEQUENCE   334 AA;  39111 MW;  96F50415 CRC32;

  Query Match             9.4%;  Score 84;  DB 1;  Length 334;
  Best Local Similarity  64.3%;  Pred. No. 1.09e+00;
  Matches     9;  Conservative    5;  Mismatches   0;  Indels   0;  Gaps   0;

Db     42 FLSVILGLLLLVSC 55
          :|:||||:|||::|
Qy     32 ILTVILGVLLLIGC 45


RESULT   15
ID   CD80_MOUSE     STANDARD;      PRT;   306 AA.
AC   Q00609;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   T LYMPHOCYTE ACTIVATION ANTIGEN CD80 PRECURSOR (ACTIVATION B7-1
DE   ANTIGEN) (B7).
GN   CD80 OR B7.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

```
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=B-CELL;
RX   MEDLINE; 91341422.
RA   GRAY G.S., FREEMAN G.J., GIMMI C.D., LOMBARD D.B., ZHOU L.J.,
RA   WHITE M., FINGEROTH J.D., GRIBBEN J.G., NADLER L.M.;
RT   "Structure, expression, and T cell costimulatory activity of the
RT   murine homologue of the human B lymphocyte activation antigen B7.";
RL   J. Exp. Med. 174:625-631(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=B-CELL;
RX   MEDLINE; 93307789.
RA   SELVAKUMAR A., WHITE P.C., DUPONT B.;
RT   "Genomic organization of the mouse B-lymphocyte activation antigen
RT   B7.";
RL   Immunogenetics 38:292-295(1993).
CC   -!- FUNCTION: INVOLVED IN THE COSTIMULATORY SIGNAL ESSENTIAL FOR T
CC       LYMPHOCYTES ACTIVATION. T CELL PROLIFERATION AND CYTOKINE
CC       PRODUCTION IS INDUCED BY THE BINDING OF CD28 OR CTLA-4 TO THIS
CC       RECEPTOR.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED ON ACTIVATED B CELLS, GAMMA
CC       INTERFERON STIMULATED MONOCYTES AND NONCIRCULATING B-CELL
CC       MALIGNANCIES.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED BETWEEN 4 AND 12 HOURS POST-
CC       ACTIVATION. PROTEIN WAS DETECTED AT CELL SURFACE AT 24 HOURS AND
CC       IT'S EXPRESSION WAS MAXIMAL FROM 48 TO 72 HOURS POST-ACTIVATION.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE C2-LIKE AND ONE V-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60958; CAA43291.1; -.
DR   EMBL; L12589; AAA37240.1; ALT_SEQ.
DR   EMBL; L12585; AAA37240.1; JOINED.
DR   EMBL; L12586; AAA37240.1; JOINED.
DR   EMBL; L12587; AAA37240.1; JOINED.
DR   EMBL; L12588; AAA37240.1; JOINED.
DR   PIR; S17291; S17291.
DR   MGD; MGI:101775; CD80.
DR   PFAM; PF00047; ig; 2.
KW   Immunoglobulin domain; T-cell; Glycoprotein; Signal; Transmembrane;
KW   Receptor.
FT   SIGNAL        1     37
FT   CHAIN        38    306       T LYMPHOCYTE ACTIVATION ANTIGEN CD80.
FT   DOMAIN       38    246       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    247    268       POTENTIAL.
FT   DOMAIN      269    306       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       47    126       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      158    226       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      227    246       IG-HINGE LIKE (POTENTIAL).
FT   DISULFID     54    119       POTENTIAL.
FT   DISULFID    165    219       POTENTIAL.
FT   CARBOHYD     93     93       POTENTIAL.
FT   CARBOHYD     99     99       POTENTIAL.
FT   CARBOHYD    149    149       POTENTIAL.
FT   CARBOHYD    189    189       POTENTIAL.
FT   CARBOHYD    210    210       POTENTIAL.
FT   CARBOHYD    214    214       POTENTIAL.
SQ   SEQUENCE   306 AA;  34589 MW;  86F8D183 CRC32;

  Query Match           9.3%;  Score 83;  DB 1;  Length 306;
  Best Local Similarity  27.6%;  Pred. No. 1.53e+00;
  Matches     8;  Conservative   14;  Mismatches   6;  Indels   1;  Gaps   1;

Db    252 GAGFGAVITVVVIVVIIKC-FCKHRSCFR 279
          :||:| :  :: ::::| | :|::|: :|
Qy     27 AAGIGILTVILGVLLLIGCWYCRRRNGYR 55


Search completed: Fri May  5 21:47:42 2000
Job time : 56 secs.
```

```
KW   Secreted protein; human; cell proliferation; cytokine activity;
KW   tissue growth; cellular differentiation; regeneration; activin;
KW   inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;
KW   anti-inflammatory activity; biomarker.
OS   Homo sapiens.
PN   WO9825959-A2.
PD   18-JUN-1998.
PF   11-DEC-1997; U22787.
PR   11-DEC-1996; US-032757.
PA   (CHIR ) CHIRON CORP.
PI   Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
DR   WPI; 98-348453/30.
DR   N-PSDB; V43602.
PT   Secreted human polypeptides - having cytokine, cell proliferation or
PT   differentiation, activin or inhibin, tumour inhibition or
PT   anti-inflammatory activities
PS   Claim 1; Pages 49-50; 78pp; English.
CC   This represents a human secreted protein. The specification provides
CC   secreted protein sequences (W63681 to W63699) encoded by the nucleic
CC   acid sequences shown in V43601 to V43619. The invention provides a
CC   method of identifying a secreted polypeptide which is modified by rough
CC   microsomes. The secreted proteins can be used in assays to determine
CC   biological activities, such as cytokine, cell proliferation, or cellular
CC   differentiation activities, tissue growth or regeneration, activin or
CC   inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC   thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC   anti-inflammatory activity.  The proteins can also be used as biomarkers,
CC   to identify tissues or cell types which express the proteins, or a stage-
CC   or disease-specific alteration in protein expression. They can be used
CC   in protein interaction assays, to identify ligands or binding proteins.
CC   Compounds which affect the biological activities of the secreted proteins
CC   or their ability to interact with specific ligands can be identified
CC   using the proteins in screening assays. The proteins and antibodies that
CC   bind specifically to the protein can also be used to design diagnostic
CC   tests and therapeutic compositions for diseases which may be associated
CC   with altered expression of these proteins. Fusion proteins comprising,
CC   e.g. signal sequences or transmembrane domains of the proteins can be
CC   used to target other protein domains to cellular membrane or they can
CC   be secreted extracellularly.
SQ   Sequence   291 AA;

  Query Match           10.6%;  Score 94;  DB 1;  Length 291;
  Best Local Similarity  53.1%;  Pred. No. 5.23e+00;
  Matches    17;  Conservative    7;  Mismatches   4;  Indels   4;  Gaps   4;

Db    257 ALAVAVLKTVILGLLCLLLL-WW-RRRKGSRA 286
          | ::::| ||||||:| ||:  |  |||:| ||
Qy     27 AAGIGIL-TVILGVL-LLIGCWYCRRRNGYRA 56


RESULT    6
ID   R05433 standard; protein; 380 AA.
AC   R05433;
DT   30-JUL-1990  (first entry)
DE   cPA-P2 Hybrid plasminogen activator.
KW   Plasminogen activator; fibrin; urokinase; thromboembolic
KW   disease; ds.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   domain          1. .23
FT                   /label=Secretory leader.
FT   domain          24. .102
FT                   /label=Plasminogen Kringle 1
FT   domain          103. .116
FT                   /label=Urokinase linker
FT   domain          117. .380
FT                   /label=ScuPA protease domain.
PN   WO9001332-A.
PD   22-FEB-1990.
PF   10-AUG-1988; 02771.
PR   10-AUG-1988; WO-U02771.
PA   (CETU) Cetus Corp.
PI   Halluin AP;
DR   WPI; 90-083374/11.
DR   N-PSDB; Q02301.
PT   Compsn. contg. plasminogen activator conjugated to heparin component -
PT   used for treatment of thromboembolic disease, with longer half
PT   life and improved targeting.
PS   Disclosure; p; English.
CC   Gene encodes hybrid plasminogen activator (PA) comprising Kringle 1, an
CC   urokinase linker, and an urokinase protease domain wherin glycine residue
CC   at position 158 is replaced with a lysine.
CC   The compound is used to treat thromboembolic disease esp. with myocardial
CC   infarction, has a longer half-life than free PA and targets the heparin
CC   site of thrombus or embolism reducing the risk of reocclusion.
SQ   Sequence   380 AA;

  Query Match           10.1%;  Score 90;  DB 1;  Length 380;
  Best Local Similarity  30.9%;  Pred. No. 1.11e+01;
  Matches    21;  Conservative   15;  Mismatches  24;  Indels   8;  Gaps   7;

Db      7 LALLLLLLLLPGCWASECKTGDGKNYRGTMSKTKN-GITCQKWSSTSPHRPRFSPATHPS 65
          |:::| :||| |||   |:  :|  ||: | |: : |  |   :   |:   |     :
Qy     33 LTVILGVLLLIGCW-Y-CRRRNG--YRALMDKSLHVGTQCAL-TRRCPQEG-FDH-RDSK 85

Db     66 EGLEENYC 73
              :|:|: |
Qy     86 VSLQEKNC 93


RESULT    7
ID   W80407 standard; Protein; 226 AA.
AC   W80407;
DT   13-JAN-1999  (first entry)
DE   A secreted protein encoded by clone di39_9.
KW   Secreted protein; immune stimulating; suppressing;
KW   haematopoiesis regulating activity; tissue growth activity; activin;
KW   inhibin activity; chemotactic; chemokinetic activity; haemostatic;
KW   thrombolytic activity; anti-inflammatory activity; cadherin;
KW   tumour invasion suppressor activity; tumour inhibition activity.
OS   Homo sapiens.
PN   WO9844113-A1.
PD   08-OCT-1998.
PF   27-MAR-1998; U06176.
PR   25-MAR-1998; US-823330.
PR   28-MAR-1997; US-823330.
PA   (GEMY ) GENETICS INST INC.
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI   Racie LA, Spaulding V, Treacy M;
DR   WPI; 98-542703/46.
DR   N-PSDB; V63191.
PT   New isolated polynucleotide(s) and secreted proteins  - are obtained
PT   from human cDNA libraries prepared from adult testes, foetal brain,
PT   adult brain, adult blood and placenta
PS   Claim 19; Pages 76-77; 124pp; English.
CC   The present sequence represents a secreted protein. The nucleic acid
CC   sequence is isolated from a human adult testes  cDNA library using
CC   probe V63202. The polypeptide may have biological activities such has
CC   e.g. nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity or other activities.
SQ   Sequence   226 AA;

  Query Match           10.0%;  Score 89;  DB 1;  Length 226;
  Best Local Similarity  53.1%;  Pred. No. 1.34e+01;
  Matches    17;  Conservative    7;  Mismatches   3;  Indels   5;  Gaps   4;

Db    193 ALAVAVLKTVILGLLCLLL--WW-RRRKGSRA 221
          | ::::| ||||||:| ||:  |  |||:| ||
Qy     27 AAGIGIL-TVILGVL-LLIGCWYCRRRNGYRA 56


RESULT    8
```

```
ID   W20924 standard; protein; 141 AA.
AC   W20924;
DT   21-JUL-1997  (first entry)
DE   H. pylori cell envelope protein, 16ae10505orf13.
KW   cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T68177.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 56; Page 1322; 1481pp; English.
CC   The present sequence is a H. pylori cell envelope protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   141 AA;

  Query Match             9.9%;  Score 88;  DB 1;  Length 141;
  Best Local Similarity  55.0%;  Pred. No. 1.61e+01;
  Matches    11;  Conservative    5;  Mismatches   4;  Indels   0;  Gaps   0;

Db    108 AEGLGITIMILGILVLLGLW 127
          | |:||  :|||:|:|:| |
Qy     27 AAGIGILTVILGVLLLIGCW 46


RESULT   9
ID   Y11007 standard; Protein; 169 AA.
AC   Y11007;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF hp6p10509_14642217_c2_17 cellular protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30474.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 217; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   169 AA;

  Query Match             9.9%;  Score 88;  DB 1;  Length 169;
  Best Local Similarity  55.0%;  Pred. No. 1.61e+01;
  Matches    11;  Conservative    5;  Mismatches   4;  Indels   0;  Gaps   0;

Db    104 AEGLGITIMILGILVLLGLW 123
          | |:||  :|||:|:|:| |
Qy     27 AAGIGILTVILGVLLLIGCW 46


RESULT   10
ID   Y11016 standard; Protein; 215 AA.
AC   Y11016;
DT   08-JUN-1999  (first entry)
DE   H. pylori ORF hp6p10509_14642217_c3_25 cellular protein.
KW   Vaccine; probe; diagnostic; ORF; cell envelope protein;
KW   secreted protein; cellular protein.
OS   Helicobacter pylori.
PN   WO9818323-A1.
PD   07-MAY-1998.
PF   28-OCT-1997; U19575.
PR   14-JUL-1997; US-891928.
PR   28-OCT-1996; US-739150.
PR   06-DEC-1996; US-759739.
PA   (ASTR ) ASTRA AB.
PI   Alm RA, Smith D;
DR   WPI; 98-271811/24.
DR   N-PSDB; X30483.
PT   Helicobacter pylori nucleic acids and proteins - used to develop
PT   products for the detection, prevention and treatment of H. pylori
PT   infections
PS   Claims 27, 31; Page 226-227; 279pp; English.
CC   Recombinant or substantially pure preparations of H. pylori polypeptides
CC   are disclosed, together with the nucleic acids encoding them. In all,
CC   73 ORFs are shown. The proteins are variously cell envelope proteins,
CC   secreted proteins or other cellular proteins. Vaccines containing the
CC   nucleic acids or proteins are claimed, as are probes containing at least
CC   8 nucleotides from the nucleic acid sequences. The vaccines are useful
CC   for treating or reducing the risk of H. pylori infections, and the
CC   probes can be used diagnostically for detecting the presence of
CC   Helicobacter in a sample. The products are also of use in screening
CC   for compounds having the ability to interfere with the H. pylori life
CC   cycle or to inhibit H. pylori infection.
SQ   Sequence   215 AA;

  Query Match             9.9%;  Score 88;  DB 1;  Length 215;
  Best Local Similarity  55.0%;  Pred. No. 1.61e+01;
  Matches    11;  Conservative    5;  Mismatches   4;  Indels   0;  Gaps   0;

Db    104 AEGLGITIMILGILVLLGLW 123
          | |:||  :|||:|:|:| |
Qy     27 AAGIGILTVILGVLLLIGCW 46


RESULT   11
ID   W30826 standard; Protein; 519 AA.
AC   W30826;
DT   20-MAR-1998  (first entry)
DE   The novel tyrosinase-related protein 2 (TRP-2).
KW   Tyrosinase related protein 2 gene; TRP-1; gp75; tumour antigen;
KW   tumour infiltrating lymphocyte; TIL; TIL586; cancer peptide; TRP-2;
KW   alternative reading frame; cancer detection; pre-cancer detection;
KW   melanoma.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         197. .205
```

```
FT                    /note= "antigenic peptide sequence recognised by CTL"
PN   WO9729195-A2.
PD   14-AUG-1997.
PF   06-FEB-1997; U02186.
PR   04-OCT-1996; US-725736.
PR   09-FEB-1996; US-599602.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI   Rosenberg SA, Wang R;
DR   WPI; 97-415349/38.
DR   N-PSDB; T91957.
PT   Cancer antigen peptide(s) derived from the tyrosinase-related
PT   protein 1 or 2  - useful for detecting, preventing or treating a
PT   cancer in a mammal, especially melanoma
PS   Claim 12; Pages 81-83; 111pp; English.
CC   The present sequence represents the novel tyrosinase related
CC   protein 2 (TRP-2). This protein contains tumour antigens recognised
CC   by tumour infiltrating lymphocyte (TIL) 586. Novel cancer peptides have
CC   also been identified in TRP-1. The peptides are recognised by a major
CC   histocompatibility complex (MHC) class I T-lymphocyte. The nucleic acids
CC   encoding the cancer peptides or TRP-2 can be used to detect a cancer or
CC   pre-cancer in a mammal, especially by detecting the presence of the
CC   alterative ORF 3 of the TRP-1 gene or the sequence encoding the novel
CC   tumour antigen TRP-2. Vectors and recombinant viruses containing
CC   antigen peptide encoding nucleic acids, antibodies raised against the
CC   peptides, or the peptides themselves can be used to prevent or treat
CC   a cancer in a mammal, especially a melanoma.
SQ   Sequence   519 AA;

  Query Match             9.8%;  Score 87;  DB 1;  Length 519;
  Best Local Similarity  35.3%;  Pred. No. 1.94e+01;
  Matches    12;  Conservative   11;  Mismatches  10;  Indels   1;  Gaps   1;

Db    477 MGTLVALVGLFVLLAFLQYRRLRKGYTPLMETHL 510
          :| |  ::|:::|::    || |:|| :||:  |
Qy     30 IGILTVILGVLLLIGCWYCRR-RNGYRALMDKSL 62


RESULT   12
ID   W20802 standard; protein; 296 AA.
AC   W20802;
DT   16-JUL-1997  (first entry)
DE   H. pylori inner membrane protein, 09ap11406orf5.
KW   Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW   binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW   duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
OS   Helicobacter pylori.
PN   WO9640893-A1.
PD   19-DEC-1996.
PF   06-JUN-1996; U09122.
PR   07-JUN-1995; US-487032.
PR   01-APR-1996; US-630405.
PA   (ASTR ) ASTRA AB.
PI   Berglindh OT,  Smith D,  Mellgaerd BL;
DR   WPI; 97-052306/05.
DR   N-PSDB; T68055.
PT   Helicobacter pylori nucleic acid sequences and related
PT   polypeptide(s)  - useful for vaccines to treat or prevent H. pylori
PT   infection, and to detect Helicobacter
PS   Claim 56; Page 1209; 1481pp; English.
CC   The present sequence is a H. pylori inner membrane protein.
CC   The protein may be used in a vaccine to prevent or treat H. pylori
CC   infection or to identify H. pylori polypeptide binding compounds,
CC   useful as potential H. pylori life cycle activators or inhibitors.
CC   The genomic sequence of H. pylori (ATCC 55679) was determined from
CC   overlapping contigs generated by mechanically shearing the bacterial
CC   DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC   and the predicted coding regions defined by computer evaluation. To
CC   identify likely H. pylori antigens for vaccine development, the amino
CC   acid sequences predicted from various ORF were analysed for significant
CC   homology to other known or exported membrane proteins. Having identified
CC   and determined the sequences of interest, particular regions can be
CC   isolated from H. pylori by PCR amplification for recombinant polypeptide
CC   production, e.g. in E. coli hosts.
SQ   Sequence   296 AA;

  Query Match             9.4%;  Score 84;  DB 1;  Length 296;
  Best Local Similarity  29.7%;  Pred. No. 3.37e+01;
  Matches    11;  Conservative   14;  Mismatches  11;  Indels   1;  Gaps   1;

Db    210 FFGNLKTGNNQISVFEDLNAREVGVLSVILALILILG 246
          ::|  | |: :     : :|  :|:|:|||:::|::|
Qy      9 IYGYPKKGHGHSYTTAE-EAAGIGILTVILGVLLLIG 44


RESULT   13
ID   R82900 standard; Protein; 200 AA.
AC   R82900;
DT   07-MAY-1996  (first entry)
DE   Mouse B7-1 (IgV-like domain deleted).
KW   T-cell costimulatory molecule; B7-1; T-lymphocyte; CD28; CTLA4;
KW   receptor; immunoglobulin.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1. .37
FT                   /label= Sig_peptide
FT   domain          169. .200
FT                   /label= Cytoplasmic_domain
FT                   /note= "cytoplasmic domain is encoded by exon 5
FT                   of the B7-1 gene"
PN   WO9523859-A2.
PD   08-SEP-1995.
PF   02-MAR-1995; U02576.
PR   02-MAR-1994; US-205697.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (DAND ) DANA FARBER CANCER INST.
PI   Borriello F,  Freeman GJ,  Nadler LM,  Sharpe AH;
DR   WPI; 95-320574/41.
DR   N-PSDB; T01047.
PT   Novel T cell co-stimulatory molecules - corresponding to naturally
PT   occurring alternatively spliced forms of T cells co-stimulatory
PT   molecules or variants
PS   Disclosure; Page 55-56; 111pp; English.
CC   A naturally occurring form of mouse T-cell costimulatory molecule
CC   B7-1 (R82900) has the signal peptide directly linked to the
CC   IgC-like domain, i.e. the IgV-like domain is deleted.  It is
CC   encoded by exons 1, 3, 4 and 5 (see T01047) of the B7-1 gene.  An
CC   alternatively spliced form of IgV-deleted B7-1 (R82901) is encoded
CC   by exons 1, 3, 4 and 6.  T-cell costimulatory molecules can be
CC   produced in which the IgV-like domain is deleted.
SQ   Sequence   200 AA;

  Query Match             9.3%;  Score 83;  DB 1;  Length 200;
  Best Local Similarity  27.6%;  Pred. No. 4.04e+01;
  Matches     8;  Conservative   14;  Mismatches   6;  Indels   1;  Gaps   1;

Db    146 GAGFGAVITVVVIVVIIKC-FCKHRSCFR 173
          :||:| :  :: ::::| | :|::|: :|
Qy     27 AAGIGILTVILGVLLLIGCWYCRRRNGYR 55


RESULT   14
ID   R82902 standard; Protein; 212 AA.
AC   R82902;
DT   07-MAY-1996  (first entry)
DE   Mouse B7-1 IgV-like isoform.
KW   T-cell costimulatory molecule; B7-1; T-lymphocyte; CD28; CTLA4;
KW   receptor; immunoglobulin; interleukin-2.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   peptide         1. .37
FT                   /label= Sig_peptide
FT   domain          181. .212
FT                   /label= Cytoplasmic_domain
FT                   /note= "cytoplasmic domain is encoded by exon 5
FT                   of the B7-1 gene"
PN   WO9523859-A2.
```

```
PD   08-SEP-1995.
PF   02-MAR-1995; U02576.
PR   02-MAR-1994; US-205697.
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL.
PA   (DAND ) DANA FARBER CANCER INST.
PI   Borriello F, Freeman GJ, Nadler LM, Sharpe AH;
DR   WPI; 95-320574/41.
DR   N-PSDB; T01049.
PT   Novel T cell co-stimulatory molecules - corresponding to naturally
PT   occurring alternatively spliced forms of T cells co-stimulatory
PT   molecules or variants
PS   Disclosure; Page 91-92; 111pp; English.
CC   A naturally occurring form of mouse T-cell costimulatory molecule
CC   B7-1 (R82902) has the IgV-like domain directly linked to the
CC   transmembrane domain, i.e. the IgC-like domain is deleted. It is
CC   encoded by exons 1, 2, 4 and 5 of the B7-1 (T01049) gene.
CC   This IgV-like isoform of B701 was expressed in CHO cells. It
CC   triggered a costimulatory signal in T-cells, causing stimulation
CC   of interleukin-2 prodn.
SQ   Sequence  212 AA;

  Query Match            9.3%;  Score 83;  DB 1;  Length 212;
  Best Local Similarity  27.6%;  Pred. No. 4.04e+01;
  Matches    8;  Conservative   14;  Mismatches   6;  Indels   1;  Gaps   1;

Db    158 GAGFGAVITVVVIVVIIKC-FCKHRSCFR 185
          :||:| :  :: ::::| | :|::|: :|
Qy     27 AAGIGILTVILGVLLLIGCWYCRRRNGYR 55


RESULT  15
ID   R67990 standard; Protein; 306 AA.
AC   R67990;
DT   21-AUG-1995  (first entry)
DE   Murine B lymphocyte antigen B7 (mB7).
KW   B lymphocyte activation antigen; B7-1; Ig superfamily; CD28;
KW   transmembrane protein.
OS   Mus musculus.
FH   Key             Location/Qualifiers
FT   protein         1. .37
FT                   /label= signal sequence
FT                   /note= "hydrophobic"
FT   domain          38. .247
FT                   /label= extracellular
FT                   /note= "6"
FT   domain          248. .272
FT                   /label= transmembrane
FT   domain          273. .306
FT                   /label= intracellular (cytoplasmic)
FT   domain          38. .142
FT                   /label= Ig V-set domain
FT   domain          143. .236
FT                   /label= Ig c-set domain
FT   misc_difference 1. .306
FT                   /label= published
FT                   /note= "Freeman, G.J. et al. see CC"
PN   WO9503408-A.
PD   02-FEB-1995.
PF   26-JUL-1994; U08423.
PR   26-JUL-1993; US-101624.
PR   19-AUG-1993; US-109393.
PR   03-NOV-1993; US-147773.
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REPK ) REPLIGEN CORP.
PI   Freeman GJ, Gray GS, Greenfield E, Nadler LM;
DR   WPI; 95-075236/10.
DR   N-PSDB; Q81372.
PT   Nucleic acids encoding CTLA4/CD28 counter receptor, B7-2 - useful
PT   for enhancing or suppressing T-cell mediated immune responses
PS   Disclosure; pages 118-120; 175pp; English.
CC   Q81371 is in pCDM8 vector. It is derived from germline B
CC   lymphocytes, cell lines 70Z and A20, clones B7 nos. 1 and 29.
CC   It can be found in Genbank at Accession no. X60958. The encoded
CC   protein, R67990, binds both human CTLA4 and human CD28. It is
CC   related to human hB7-2 (see Q81351) and human hB7-1 (see Q81371).
CC   Part of R67990 (see CC) is published in Freeman, G.J. et al.
CC   J. of Experimental Medicine, in press at the time when the patent
CC   application was written.
SQ   Sequence  306 AA;

  Query Match            9.3%;  Score 83;  DB 1;  Length 306;
  Best Local Similarity  27.6%;  Pred. No. 4.04e+01;
  Matches    8;  Conservative   14;  Mismatches   6;  Indels   1;  Gaps   1;

Db    252 GAGFGAVITVVVIVVIIKC-FCKHRSCFR 279
          :||:| :  :: ::::| | :|::|: :|
Qy     27 AAGIGILTVILGVLLLIGCWYCRRRNGYR 55


Search completed: Fri May  5 21:44:20 2000
Job time : 51 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:01:50 2000; MasPar time 7.16 Seconds
87.109 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-4
Description: (1-9) from US09267439.pep
Perfect Score: 56
Sequence: 1 AAGIGILTV 9

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl12
1:sp_archea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.729; Variance 25.002; scale 0.869

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	49	87.5	766	10	O23161	RECEPTOR KINASE-LIKE P	3.66e+00
2	48	85.7	808	10	Q9ZT37	PUTATIVE GLUTAMATE REC	6.11e+00
3	47	83.9	165	2	P71810	HYPOTHETICAL 18.2 KD P	1.01e+01
4	47	83.9	250	2	O31597	YJBA PROTEIN.	1.01e+01
5	47	83.9	478	14	Q87090	GLYCOPROTEIN GIII.	1.01e+01
6	47	83.9	479	14	Q87089	GLYCOPROTEIN GIII.	1.01e+01
7	47	83.9	479	14	Q87091	GLYCOPROTEIN GIII.	1.01e+01
8	46	82.1	848	5	O18139	T26H2.7 PROTEIN.	1.67e+01
9	45	80.4	339	1	O30640	METHYLCOBAMIDE:COM MET	2.73e+01
10	45	80.4	339	1	Q48928	METHYLCOBAMIDE:COM MET	2.73e+01
11	45	80.4	339	1	Q48950	METHYLCOBALAMIN: COENZ	2.73e+01
12	45	80.4	370	8	O48172	CYTOCHROME B.	2.73e+01
13	45	80.4	420	11	Q63276	KAN-1.	2.73e+01
14	45	80.4	420	11	O08833	BILE ACID COA: AMINO A	2.73e+01
15	45	80.4	980	5	Q17592	SIMILARITY TO INSULIN-	2.73e+01
16	45	80.4	1347	2	O30426	XYLANASE.	2.73e+01
17	44	78.6	98	2	Q9Z8X3	HYPOTHETICAL 10.3 KD P	4.43e+01
18	44	78.6	620	1	O29198	ABC TRANSPORTER, ATP-B	4.43e+01
19	43	76.8	91	11	O54712	ACETYLCHOLINE RECEPTOR	7.14e+01
20	43	76.8	190	10	Q9XIA3	F13F21.20 PROTEIN.	7.14e+01
21	43	76.8	243	2	O86207	PHOSPHOGLYCOLATE PHOSP	7.14e+01
22	43	76.8	361	2	Q9X336	PXO1-66.	7.14e+01
23	43	76.8	509	2	O05457	HYPOTHETICAL 53.2 KD P	7.14e+01
24	43	76.8	520	5	P91840	W05H5.3 PROTEIN.	7.14e+01
25	43	76.8	579	2	Q9X7A9	PUTATIVE ACYLTRANSFERA	7.14e+01
26	43	76.8	580	2	O53208	PUTATIVE TRANSFERASE.	7.14e+01
27	43	76.8	593	2	P94513	AUTOLYSIN SENSOR KINAS	7.14e+01
28	43	76.8	667	2	P71749	HYPOTHETICAL 68.3 KD P	7.14e+01
29	43	76.8	773	10	O22178	PUTATIVE RECEPTOR PROT	7.14e+01
30	43	76.8	7576	2	Q9ZGA4	FK506 POLYKETIDE SYNTH	7.14e+01
31	42	75.0	12	2	Q46039	AMPR GENE	1.14e+02
32	42	75.0	116	2	Q9XDM7	PDUU.	1.14e+02
33	42	75.0	264	2	O83139	CYTOCHROME C BIOGENESI	1.14e+02
34	42	75.0	303	2	O06279	HYPOTHETICAL 31.1 KD P	1.14e+02
35	42	75.0	346	10	Q9ZQ87	T4M8.5 PROTEIN.	1.14e+02
36	42	75.0	365	2	Q9X0A4	CONSERVED HYPOTHETICAL	1.14e+02
37	42	75.0	406	2	P77444	FROM BASES 1755151 TO	1.14e+02
38	42	75.0	417	2	O66600	SULFIDE DEHYDROGENASE,	1.14e+02
39	42	75.0	446	2	Q9ZBV5	PUTATIVE INTEGRAL MEMB	1.14e+02
40	42	75.0	464	2	O67342	NADH DEHYDROGENASE I C	1.14e+02
41	42	75.0	502	5	Q22770	T25B9.7 PROTEIN.	1.14e+02
42	42	75.0	503	2	O26074	PROTEIN-EXPORT MEMBRAN	1.14e+02
43	42	75.0	509	2	O85163	MULTIDRUG RESISTANCE E	1.14e+02
44	42	75.0	526	2	Q9ZJ66	PROTEIN-EXPORT MEMBRAN	1.14e+02
45	42	75.0	596	5	Q18472	C35A5.3 PROTEIN.	1.14e+02

ALIGNMENTS

RESULT 1
ID O23161 PRELIMINARY; PRT; 766 AA.
AC O23161;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE RECEPTOR KINASE-LIKE PROTEIN (EC 2.7.1.).
GN C7A10.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA BEVAN M., TERRYN N., VOS P., HEIJNEN L., MEWES H.W., SCHUELLER C.,
RA CHALWATZIS N.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z99707; CAB16774.1; -.
DR MENDEL; 25486; Arath;3435;25486.
DR PFAM; PF00560; LRR; 4.
DR PFAM; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEURICHRPT.
SQ SEQUENCE 766 AA; 83775 MW; C4BD7115 CRC32;

Query Match 87.5%; Score 49; DB 10; Length 766;
Best Local Similarity 87.5%; Pred. No. 3.66e+00;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 337 AGIGILAV 344
||||||:|
Qy 2 AGIGILTV 9

RESULT 2
ID Q9ZT37 PRELIMINARY; PRT; 808 AA.
AC Q9ZT37;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PUTATIVE GLUTAMATE RECEPTOR.
GN GLR1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 99039497.
RA   LAM H.M., CHIU J., HSIEH M.H., MEISEL L., OLIVEIRA I.C., SHIN M.,
RA   CORUZZI G.;
RT   "Glutamate-receptor genes in plants.";
RL   Nature 396:125-126(1998).
DR   EMBL; AF079998; AAD09173.1; -.
KW   Receptor.
SQ   SEQUENCE   808 AA;  90518 MW;  C3554B89 CRC32;

  Query Match          85.7%;  Score 48;  DB 10;  Length 808;
  Best Local Similarity 100.0%;  Pred. No. 6.11e+00;
  Matches    7;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db    519 GIGILTV 525
          |||||||
Qy      3 GIGILTV 9


RESULT    3
ID   P71810      PRELIMINARY;      PRT;   165 AA.
AC   P71810;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HYPOTHETICAL 18.2 KD PROTEIN.
GN   MTCY02B12.16.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   MCLEAN J., HARRIS D.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RA   BARRELL B.G., RAJANDREAM M.A.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE; 96181548.
RA   PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA   BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA   COLE S.T.;
RT   "An integrated map of the genome of the tubercle bacillus,
RT   Mycobacterium tuberculosis H37Rv, and comparison with Mycobacterium
RT   leprae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).
DR   EMBL; Z81011; CAB02643.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   165 AA;  18189 MW;  BFB84C79 CRC32;

  Query Match          83.9%;  Score 47;  DB 2;  Length 165;
  Best Local Similarity 75.0%;  Pred. No. 1.01e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    128 AGIGILAI 135
          ||||||::
Qy      2 AGIGILTV 9


RESULT    4
ID   O31597      PRELIMINARY;      PRT;   250 AA.
AC   O31597;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YJBA PROTEIN.
GN   YJBA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE; 98044033.
RA   KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA   AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA   BORRISS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA   BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA   CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA   DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA   ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA   FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA   GHIM S.Y., GLASER P., GOFFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA   GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA   HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA   JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA   KOBAYASHI Y., KOETTER P., KONINGSTEIN G., KROGH S., KUMANO M.,
RA   KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA   LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA   MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA   NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA   PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
RA   PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA   RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA   SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA   SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDO B.,
RA   SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA   TAKEUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA   TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA   VIARI A., WAMBUTT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA   WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA   YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z99110; CAB12998.1; -.
SQ   SEQUENCE   250 AA;  30119 MW;  C96222FD CRC32;

  Query Match          83.9%;  Score 47;  DB 2;  Length 250;
  Best Local Similarity 66.7%;  Pred. No. 1.01e+01;
  Matches    6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db     89 TDGIGILAV 97
          : |||||:|
Qy      1 AAGIGILTV 9


RESULT    5
ID   Q87090      PRELIMINARY;      PRT;   478 AA.
AC   Q87090;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIANA S;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
```

```
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49436; BAA08414.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   478 AA;  51150 MW;  D6A143B4 CRC32;

  Query Match           83.9%;  Score 47;  DB 14;  Length 478;
  Best Local Similarity  75.0%;  Pred. No. 1.01e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    455 AGIGILAI 462
          ||||||::
Qy      2 AGIGILTV 9


RESULT    6
ID   Q87089      PRELIMINARY;      PRT;   479 AA.
AC   Q87089;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YAMAGATA S-81;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49435; BAA08413.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   479 AA;  51109 MW;  A009EB9B CRC32;

  Query Match           83.9%;  Score 47;  DB 14;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 1.01e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      2 AGIGILTV 9


RESULT    7
ID   Q87091      PRELIMINARY;      PRT;   479 AA.
AC   Q87091;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   GLYCOPROTEIN GIII.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIA3;
RX   MEDLINE; 96316347.
RA   ISHIKAWA K., TSUTSUI M., TAGUCHI K., SAITOH A., MURAMATSU M.;
RT   "Sequence variation of the gC gene among pseudorabies virus strains.";
RL   Vet. Microbiol. 49:267-272(1996).
DR   EMBL; D49437; BAA08415.1; -.
DR   PRINTS; PR00668; GLYCPROTEINC.
SQ   SEQUENCE   479 AA;  51148 MW;  CC3EFF9A CRC32;

  Query Match           83.9%;  Score 47;  DB 14;  Length 479;
  Best Local Similarity  75.0%;  Pred. No. 1.01e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      2 AGIGILTV 9


RESULT    8
ID   O18139      PRELIMINARY;      PRT;   848 AA.
AC   O18139;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   T26H2.7 PROTEIN.
GN   T26H2.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   MATTHEWS L.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z82055; CAB04848.1; -.
SQ   SEQUENCE   848 AA;  98312 MW;  371853A7 CRC32;

  Query Match           82.1%;  Score 46;  DB 5;  Length 848;
  Best Local Similarity  66.7%;  Pred. No. 1.67e+01;
  Matches    6;  Conservative    3;  Mismatches   0;  Indels   0;  Gaps   0;

Db    340 SASIGILTI 348
          :|:||||:
Qy      1 AAGIGILTV 9


RESULT    9
ID   O30640      PRELIMINARY;      PRT;   339 AA.
AC   O30640;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METHYLCOBAMIDE:COM METHYLTRANSFERASE ISOZYME A.
GN   MTBA.
OS   Methanosarcina barkeri.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MS (DSM 800);
RX   MEDLINE; 97341199.
RA   BURKE S.A., KRZYCKI J.A.;
RT   "Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a
RT   corrinoid protein and two methyltransferases purified from
RT   Methanosarcina barkeri.";
RL   J. Biol. Chem. 272:16570-16577(1997).
DR   EMBL; AF013713; AAC38632.1; -.
DR   PFAM; PF01208; URO-D; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   339 AA;  36664 MW;  040E3CF3 CRC32;

  Query Match           80.4%;  Score 45;  DB 1;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 2.73e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;
```

```
Db    307 AGVGLLTV 314
          ||:|:|||
Qy      2 AGIGILTV 9


RESULT   10
ID   Q48928      PRELIMINARY;      PRT;   339 AA.
AC   Q48928;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METHYLCOBAMIDE:COM METHYLTRANSFERASE ISOZYME A.
GN   CMTA.
OS   Methanosarcina barkeri.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH;
RX   MEDLINE; 96324952.
RA   LECLERC G.M., GRAHAME D.A.;
RT   "Methylcobamide:coenzyme M methyltransferase isozymes from
RT   Methanosarcina barkeri. Physicochemical characterization, cloning,
RT   sequence analysis, and heterologous gene expression.";
RL   J. Biol. Chem. 271:18725-18731(1996).
DR   EMBL; U38919; AAC44214.1; -.
DR   PFAM; PF01208; URO-D; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   339 AA;  36708 MW;  731F945B CRC32;

  Query Match           80.4%;  Score 45;  DB 1;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 2.73e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      2 AGIGILTV 9


RESULT   11
ID   Q48950      PRELIMINARY;      PRT;   339 AA.
AC   Q48950;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   METHYLCOBALAMIN: COENZYME M METHYLTRANSFERASE (ISOENZYME II).
GN   MTBA.
OS   Methanosarcina barkeri.
OC   Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
OC   Methanosarcina.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FUSARO (DSM 804);
RX   MEDLINE; 96184544.
RA   HARMS U., THAUER R.K.;
RT   "Methylcobalamin: coenzyme M methyltransferase isoenzymes MtaA and
RT   MtbA from Methanosarcina barkeri. Cloning, sequencing and differential
RT   transcription of the encoding genes, and functional overexpression of
RT   the mtaA gene in Escherichia coli.";
RL   Eur. J. Biochem. 235:653-659(1996).
DR   EMBL; X91894; CAA62996.1; -.
DR   PFAM; PF01208; URO-D; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   339 AA;  36761 MW;  5F6F0A9C CRC32;

  Query Match           80.4%;  Score 45;  DB 1;  Length 339;
  Best Local Similarity  75.0%;  Pred. No. 2.73e+01;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    307 AGVGLLTV 314
          ||:|:|||
Qy      2 AGIGILTV 9


RESULT   12
ID   O48172      PRELIMINARY;      PRT;   370 AA.
AC   O48172;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   CYTOCHROME B.
GN   COB.
OS   Polytomella sp. 'Pringsheim 198.80'.
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC   Chlamydomonadaceae; Polytomella.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=198.80, FROM E.G. PRINGSHEIM;
RA   ANTARAMIAN A., FUNES-ARGUELLO S., VAZQUEZ-ACEVEDO M., CORIA R.,
RA   GONZALEZ-HALPHEN D.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U87396; AAC24896.1; -.
DR   MENDEL; 23585; Pols;cob;23585.
DR   PFAM; PF00032; cytochrome_b_C; 1.
DR   PFAM; PF00033; cytochrome_b_N; 1.
KW   Mitochondrion.
SQ   SEQUENCE   370 AA;  41226 MW;  5D617081 CRC32;

  Query Match           80.4%;  Score 45;  DB 8;  Length 370;
  Best Local Similarity  85.7%;  Pred. No. 2.73e+01;
  Matches    6;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db    288 GIGILAV 294
          |||||:|
Qy      3 GIGILTV 9


RESULT   13
ID   Q63276      PRELIMINARY;      PRT;   420 AA.
AC   Q63276;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE   KAN-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTER; TISSUE=LIVER;
RX   MEDLINE; 96003917.
RA   FURUTANI M., ARII S., HIGASHITSUJI H., MISE M., FUKUMOTO M.,
RA   TAKANO S., NAKAYAMA H., IMAMURA M., FUJITA J.;
RT   "Reduced expression of kan-1 (encoding putative bile acid-CoA-amino
RT   acid N-acyltransferase) mRNA in livers of rats after partial
RT   hepatectomy and during sepsis.";
RL   Biochem. J. 311:203-208(1995).
DR   EMBL; D43964; BAA07901.1; -.
SQ   SEQUENCE   420 AA;  46496 MW;  7B62AACF CRC32;

  Query Match           80.4%;  Score 45;  DB 11;  Length 420;
  Best Local Similarity  55.6%;  Pred. No. 2.73e+01;
  Matches    5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    226 GPGVGILSV 234
          ::|:|||:|
Qy      1 AAGIGILTV 9


RESULT   14
ID   O08833      PRELIMINARY;      PRT;   420 AA.
AC   O08833;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
```

```
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   BILE ACID COA: AMINO ACID N-ACYLTRANSFERASE.
GN   BAAT.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   FALANY C.N., FORTINBERRY H., LEITER E.H., BARNES S.;
RL   J. Lipid Res. 0:0-0(0).
DR   EMBL; U95215; AAB58325.1; -.
DR   MGD; MGI:106642; Baat.
KW   Transferase; Acyltransferase.
SQ   SEQUENCE   420 AA;  46528 MW;  4A22FFFC CRC32;

  Query Match           80.4%;  Score 45;  DB 11;  Length 420;
  Best Local Similarity  55.6%;  Pred. No. 2.73e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    225 GPGVGILSV 233
          ::|:|||:|
Qy      1 AAGIGILTV 9


RESULT   15
ID   Q17592      PRELIMINARY;      PRT;   980 AA.
AC   Q17592;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   SIMILARITY TO INSULIN-DEGRADING ENZYMES.
GN   C02G6.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC   Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE; 94150718.
RA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA   GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA   LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA   SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA   THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA   WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   BENTLEY D., KEMP K., SCHEET P.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   WATERSTON R.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U55372; AAA98001.1; -.
DR   PROSITE; PS00143; INSULINASE; 1.
DR   PFAM; PF00675; Peptidase_M16; 1.
SQ   SEQUENCE   980 AA;  112806 MW;  6D56C08D CRC32;

  Query Match           80.4%;  Score 45;  DB 5;  Length 980;
  Best Local Similarity  77.8%;  Pred. No. 2.73e+01;
  Matches     7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    303 AAGFGILNV 311
          |||:||| |
Qy      1 AAGIGILTV 9


Search completed: Fri May  5 22:03:19 2000
Job time : 89 secs.
```

****

MPSRCH (TM)

****

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 22:00:53 2000; MasPar time 3.16 Seconds
85.040 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-4
Description: (1-9) from US09267439.pep
Perfect Score: 56
Sequence: 1 AAGIGILTV 9

Scoring table: PAM 150
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 22.670; Variance 23.392; scale 0.969

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	56	100.0	118	1	MAR1_HUMAN	MELANOMA ANTIGEN RECOG	1.80e-02
2	47	83.9	101	1	ATPL_SULAC	MEMBRANE-ASSOCIATED AT	3.15e+00
3	47	83.9	479	1	VGLC_PRVIF	GLYCOPROTEIN GIII PREC	3.15e+00
4	46	82.1	231	1	YSA1_YEAST	YSA1 PROTEIN.	5.37e+00
5	45	80.4	291	1	Y4TQ_RHISN	PROBABLE PEPTIDE ABC T	9.09e+00
6	44	78.6	394	1	FTSZ_AZOVI	CELL DIVISION PROTEIN	1.52e+01
7	44	78.6	456	1	GLMU_ECOLI	UDP-N-ACETYLGLUCOSAMIN	1.52e+01
8	44	78.6	635	1	XYND_PAEPO	ENDO-1,4-BETA-XYLANASE	1.52e+01
9	43	76.8	132	1	ATPE_ARATH	ATP SYNTHASE EPSILON C	2.53e+01
10	43	76.8	404	1	SGAA_HYPME	SERINE--GLYOXYLATE AMI	2.53e+01
11	43	76.8	461	1	YXCC_BACSU	HYPOTHETICAL METABOLIT	2.53e+01
12	43	76.8	493	1	ACHE_MOUSE	ACETYLCHOLINE RECEPTOR	2.53e+01
13	43	76.8	611	1	YD3M_HERAU	HYPOTHETICAL 68.4 KD P	2.53e+01
14	42	75.0	110	1	VCAD_LAMBD	HEAD DECORATION PROTEI	4.15e+01
15	42	75.0	216	1	FLA2_METVO	FLAGELLIN B2 PRECURSOR	4.15e+01
16	42	75.0	218	1	FLA1_METVO	FLAGELLIN B1 PRECURSOR	4.15e+01
17	42	75.0	222	1	FLA2_METVA	FLAGELLIN B2 PRECURSOR	4.15e+01
18	42	75.0	308	1	MENA_HAEIN	1,4-DIHYDROXY-2-NAPHTH	4.15e+01
19	42	75.0	325	1	RCEM_CHRVI	REACTION CENTER PROTEI	4.15e+01
20	42	75.0	332	1	ACOA_ALCEU	ACETOIN:2,6-DICHLOROPH	4.15e+01
21	42	75.0	345	1	HRCA_STRMU	HEAT-INDUCIBLE TRANSCR	4.15e+01
22	42	75.0	461	1	THDF_HAEIN	POSSIBLE THIOPHENE AND	4.15e+01
23	42	75.0	501	1	LYSI_CORGL	L-LYSINE TRANSPORT PRO	4.15e+01
24	42	75.0	503	1	SECD_HELPY	PROTEIN-EXPORT MEMBRAN	4.15e+01
25	42	75.0	526	1	SECD_HELPJ	PROTEIN-EXPORT MEMBRAN	4.15e+01
26	42	75.0	530	1	AIP2_YEAST	ACTIN INTERACTING PROT	4.15e+01
27	42	75.0	659	1	YYBT_BACSU	HYPOTHETICAL 74.3 KD P	4.15e+01
28	42	75.0	885	1	YDGH_BACSU	PUTATIVE MEMBRANE PROT	4.15e+01
29	42	75.0	1325	1	YDEK_ECOLI	HYPOTHETICAL 136.5 KD	4.15e+01
30	42	75.0	1530	1	BFR1_SCHPO	BREFELDIN A RESISTANCE	4.15e+01
31	41	73.2	225	1	CD9_FELCA	CD9 ANTIGEN.	6.75e+01
32	41	73.2	237	1	BACT_HALSA	SENSORY RHODOPSIN II (	6.75e+01
33	41	73.2	271	1	YK23_YEAST	HYPOTHETICAL 31.0 KD P	6.75e+01
34	41	73.2	313	1	Y4TP_RHISN	PROBABLE PEPTIDE ABC T	6.75e+01
35	41	73.2	337	1	OPSX_HUMAN	VISUAL PIGMENT-LIKE RE	6.75e+01
36	41	73.2	359	1	YFDA_CORGL	HYPOTHETICAL PROTEIN I	6.75e+01
37	41	73.2	384	1	PQQE_METEX	COENZYME PQQ SYNTHESIS	6.75e+01
38	41	73.2	401	1	YABA_SCHPO	HYPOTHETICAL 44.4 KD P	6.75e+01
39	41	73.2	487	1	Y346_MYCTU	HYPOTHETICAL 52.2 KD T	6.75e+01
40	41	73.2	633	1	Y561_HAEIN	HYPOTHETICAL PROTEIN H	6.75e+01
41	41	73.2	666	1	CO14_BRAJA	PROBABLE CYTOCHROME C	6.75e+01
42	41	73.2	845	1	MAT3_RAT	MATRIN 3.	6.75e+01
43	41	73.2	977	1	YDG8_SCHPO	HYPOTHETICAL 111.4 KD	6.75e+01
44	41	73.2	1109	1	CYGD_CANFA	RETINAL GUANYLYL CYCLA	6.75e+01
45	41	73.2	1331	1	CYAB_LEIDO	RECEPTOR-TYPE ADENYLAT	6.75e+01

ALIGNMENTS

```
RESULT    1
ID   MAR1_HUMAN     STANDARD;      PRT;   118 AA.
AC   Q16655;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MELANOMA ANTIGEN RECOGNIZED BY T-CELLS 1 (MART-1) (MELAN-A PROTEIN)
DE   (ANTIGEN SK29-AA) (ANTIGEN LB39-AA).
GN   MLANA OR MART1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RX   MEDLINE; 94224770.
RA   KAWAKAMI Y., ELIYAHU S., DELGADO C.H., ROBBINS P.F., RIVOLTINI L.,
RA   TOPALIAN S.L., MIKI T., ROSENBERG S.A.;
RT   "Cloning of the gene coding for a shared human melanoma antigen
RT   recognized by autologous T cells infiltrating into tumor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:3515-3519(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94275389.
RA   COULIE P.G., BRICHARD V., VAN PEL A., WOELFEL T., SCHNEIDER J.,
RA   TRAVERSARI C., MATTEI S., DE PLAEN E., LURQUIN C., SZIKORA J.-P.,
RA   RENAULD J.-C., BOON T.;
RT   "A new gene coding for a differentiation antigen recognized by
RT   autologous cytolytic T lymphocytes on HLA-A2 melanomas.";
RL   J. Exp. Med. 180:35-42(1994).
CC   -!- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO MELANOMA AND
CC       MELANOCYTE CELL LINES AND RETINA.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U06452; AAA19238.1; -.
DR   EMBL; U06654; AAA20389.1; -.
KW   Antigen; Transmembrane.
FT   TRANSMEM     27     47       POTENTIAL.
SQ   SEQUENCE   118 AA;  13157 MW;  DFE2CFC6 CRC32;
```

```
  Query Match          100.0%;  Score 56;  DB 1;  Length 118;
  Best Local Similarity 100.0%;  Pred. No. 1.80e-02;
  Matches    9;  Conservative    0;  Mismatches   0;  Indels   0;  Gaps   0;

Db     27 AAGIGILTV 35
          |||||||||
Qy      1 AAGIGILTV 9


RESULT    2
ID   ATPL_SULAC     STANDARD;      PRT;   101 AA.
AC   P23040;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   MEMBRANE-ASSOCIATED ATPASE C CHAIN (EC 3.6.1.34) (SUL-ATPASE
DE   PROTEOLIPID CHAIN).
GN   ATPP.
OS   Sulfolobus acidocaldarius.
OC   Archaea; Crenarchaeota; Sulfolobales; Sulfolobus.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE; 89214142.
RA   DENDA K., KONISHI J., OSHIMA T., DATE T., YOSHIDA M.;
RT   "A gene encoding the proteolipid subunit of Sulfolobus acidocaldarius
RT   ATPase complex.";
RL   J. Biol. Chem. 264:7119-7121(1989).
CC   -!- FUNCTION: THE C CHAIN IS A PROTEOLIPID, AND ONE OF THE MEMBRANOUS
CC       SUBUNITS OF THE THE NONENZYMATIC COMPONENT OF THE SUL-ATPASE
CC       COMPLEX.
CC   -!- SUBUNIT: SUL-ATPASE IS COMPOSED OF SIX (OR FIVE ?) SUBUNITS:
CC       ALPHA, BETA, DELTA, GAMMA, C (PROTEOLIPID), AND POSSIBLY EPSILON.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE V-ATPASE PROTEOLIPID SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04740; AAA72703.1; -.
DR   PIR; A33351; A33351.
DR   HSSP; P00138; 1CGN.
DR   PFAM; PF00137; ATP-synt_C; 1.
KW   Hydrogen ion transport; Lipid-binding; Transmembrane.
FT   TRANSMEM      5     25       POTENTIAL.
FT   TRANSMEM     37     57       POTENTIAL.
FT   TRANSMEM     75     95       POTENTIAL.
SQ   SEQUENCE   101 AA;  10362 MW;  1DC8C74D CRC32;

  Query Match          83.9%;  Score 47;  DB 1;  Length 101;
  Best Local Similarity 87.5%;  Pred. No. 3.15e+00;
  Matches    7;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     59 AAGIGVLT 66
          |||||:||
Qy      1 AAGIGILT 8


RESULT    3
ID   VGLC_PRVIF     STANDARD;      PRT;   479 AA.
AC   P06024;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   01-APR-1993 (Rel. 25, Last annotation update)
DE   GLYCOPROTEIN GIII PRECURSOR.
OS   Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 86200375.
RA   ROBBINS A.K., WATSON R.J., WHEALY M.E., HAYS W.W., ENQUIST L.W.;
RT   "Characterization of a pseudorabies virus glycoprotein gene with
RT   homology to herpes simplex virus type 1 and type 2 glycoprotein C.";
RL   J. Virol. 58:339-347(1986).
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN C.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN GENE SUPERFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M12778; AAA47464.1; -.
DR   PIR; A26097; VGBEPB.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    479       GLYCOPROTEIN GIII.
FT   CARBOHYD     40     40       POTENTIAL.
FT   CARBOHYD     84     84       POTENTIAL.
FT   CARBOHYD    169    169       POTENTIAL.
FT   CARBOHYD    192    192       POTENTIAL.
FT   CARBOHYD    220    220       POTENTIAL.
FT   CARBOHYD    228    228       POTENTIAL.
FT   CARBOHYD    285    285       POTENTIAL.
FT   CARBOHYD    302    302       POTENTIAL.
SQ   SEQUENCE   479 AA;  51206 MW;  42EE5703 CRC32;

  Query Match          83.9%;  Score 47;  DB 1;  Length 479;
  Best Local Similarity 75.0%;  Pred. No. 3.15e+00;
  Matches    6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    456 AGIGILAI 463
          ||||||::
Qy      2 AGIGILTV 9


RESULT    4
ID   YSA1_YEAST     STANDARD;      PRT;   231 AA.
AC   Q01976;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   YSA1 PROTEIN.
GN   YSA1 OR YBR111C OR YBR0907.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 95208357.
RA   MANNHAUPT G., STUCKA R., EHNLE S., VETTER I., FELDMANN H.;
RT   "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
RL   Yeast 10:1363-1381(1994).
RN   [2]
RP   SEQUENCE OF 1-47 FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE; 92327848.
RA   MANNHAUPT G., STUCKA R., EHNLE S., VETTER I., FELDMANN H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
CC   -!- SIMILARITY: STRONG, TO B.SUBTILIS YQKG.
CC   -!- SIMILARITY: TO PROTEINS WITH A CORE MUTT DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z35980; CAA85068.1; -.
DR   EMBL; X78993; CAA55614.1; -.
DR   EMBL; X66247; CAA46972.1; -.
DR   PIR; S44691; S44691.
DR   SGD; L0002551; YSA1.
DR   PROSITE; PS00893; MUTT; 1.
DR   PFAM; PF00293; mutT; 1.
FT   DOMAIN      112    145       MUTT-LIKE.
SQ   SEQUENCE   231 AA;  26087 MW;  49A2D6CB CRC32;

  Query Match           82.1%;  Score 46;  DB 1;  Length 231;
  Best Local Similarity  85.7%;  Pred. No. 5.37e+00;
  Matches    6;  Conservative    1;  Mismatches   0;  Indels   0;  Gaps   0;

Db     79 GIGILTI 85
          ||||||:
Qy      3 GIGILTV 9


RESULT    5
ID   Y4TQ_RHISN     STANDARD;      PRT;   291 AA.
AC   Q53192;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   PROBABLE PEPTIDE ABC TRANSPORTER PERMEASE PROTEIN Y4TQ.
GN   Y4TQ.
OS   Rhizobium sp. (strain NGR234).
OG   Plasmid sym pNGR234a.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97305956.
RA   FREIBERG C.A., FELLAY R., BAIROCH A., BROUGHTON W.J., ROSENTHAL A.,
RA   PERRET X.;
RT   "Molecular basis of symbiosis between Rhizobium and legumes.";
RL   Nature 387:394-401(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 96389014.
RA   FREIBERG C., PERRET X., BROUGHTON W.J., ROSENTHAL A.;
RT   "Sequencing the 500-kb GC-rich symbiotic replicon of Rhizobium sp.
RT   NGR234 using dye terminators and a thermostable 'sequenase': a
RT   beginning.";
RL   Genome Res. 6:590-600(1996).
CC   -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC       SYSTEM Y4TOPQRS FOR A PEPTIDE. PROBABLY RESPONSIBLE FOR THE
CC       TRANSLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC       PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPPBC
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z68203; CAA92399.1; -.
DR   EMBL; AE000098; AAB91870.1; -.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
DR   PFAM; PF00528; BPD_transp; 1.
KW   Hypothetical protein; Transport; Amino-acid transport; Transmembrane;
KW   Inner membrane; Plasmid.
FT   TRANSMEM     28     48       POTENTIAL.
FT   TRANSMEM     92    112       POTENTIAL.
FT   TRANSMEM    137    157       POTENTIAL.
FT   TRANSMEM    213    233       POTENTIAL.
FT   TRANSMEM    249    269       POTENTIAL.
SQ   SEQUENCE   291 AA;  30910 MW;  3263271E CRC32;

  Query Match           80.4%;  Score 45;  DB 1;  Length 291;
  Best Local Similarity  66.7%;  Pred. No. 9.09e+00;
  Matches    6;  Conservative    2;  Mismatches   1;  Indels   0;  Gaps   0;

Db    147 GPGIGILIV 155
          ::|||||  |
Qy      1 AAGIGILTV 9


RESULT    6
ID   FTSZ_AZOVI     STANDARD;      PRT;   394 AA.
AC   P77817;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   CELL DIVISION PROTEIN FTSZ.
GN   FTSZ.
OS   Azotobacter vinelandii.
OC   Bacteria; Proteobacteria; gamma subdivision; Azotobacteraceae;
OC   Azotobacter.
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=DJ116;
RX   MEDLINE; 98267010.
RA   LU C., STRICKER J., ERICKSON H.P.;
RT   "FtsZ from Escherichia coli, Azotobacter vinelandii, and Thermotoga
RT   maritima -- quantitation, GTP hydrolysis, and assembly.";
RL   Cell Motil. Cytoskeleton 40:71-86(1998).
CC   -!- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.
CC       ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF
CC       THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,
CC       AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPTATION TO
CC       BEGIN. BINDS TO AND HYDROLYZES GTP.
CC   -!- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSEMBLE AT THE INNER SURFACE
CC       OF THE CYTOPLASMIC MEMBRANE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FTSZ FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U65939; AAC24603.1; -.
DR   HSSP; Q57816; 1FSZ.
DR   PROSITE; PS01134; FTSZ_1; 1.
DR   PROSITE; PS01135; FTSZ_2; 1.
KW   Cell division; Septation; GTP-binding.
FT   NP_BIND     104    112       GTP (POTENTIAL).
SQ   SEQUENCE   394 AA;  41153 MW;  4E887134 CRC32;

  Query Match           78.6%;  Score 44;  DB 1;  Length 394;
  Best Local Similarity  77.8%;  Pred. No. 1.52e+01;
  Matches    7;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db    121 AKGLGILTV 129
          | |:|||||
Qy      1 AAGIGILTV 9


RESULT    7
```

```
ID   GLMU_ECOLI     STANDARD;      PRT;   456 AA.
AC   P17114; P76746;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE PYROPHOSPHORYLASE (EC 2.7.7.23) (N-
DE   ACETYLGLUCOSAMINE-1-PHOSPHATE URIDYLTRANSFERASE).
GN   GLMU.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 85121806.
RA   WALKER J.E., GAY N.J., SARASTE M., EBERLE A.N.;
RT   "DNA sequence around the Escherichia coli unc operon. Completion of
RT   the sequence of a 17 kilobase segment containing asnA, oriC, unc,
RT   glmS and phoS.";
RL   Biochem. J. 224:799-815(1984).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 93315143.
RA   BURLAND V.D., PLUNKETT G. III, DANIELS D.L., BLATTNER F.R.;
RT   "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT   genome: organizational symmetry around the origin of replication.";
RL   Genomics 16:551-561(1993).
RN   [3]
RP   REVISIONS.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE; 97426617.
RA   BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA   RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA   GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA   MAU B., SHAO Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   IDENTIFICATION.
RX   MEDLINE; 94012475.
RA   MENGIN-LECREULX D., VAN HEIJENOORT J.;
RT   "Identification of the glmU gene encoding N-acetylglucosamine-1-
RT   phosphate uridyltransferase in Escherichia coli.";
RL   J. Bacteriol. 175:6150-6157(1993).
CC   -!- FUNCTION: BIFUNCTIONAL ENZYME RESPONSIBLE FOR THE ACETYLATION OF
CC       GLC-N-1-P TO GIVE GLCNAC-1-P AND THE SYNTHESIS OF UDP-GLCNAC.
CC   -!- CATALYTIC ACTIVITY: UTP + N-ACETYL-ALPHA-D-GLUCOSAMINE
CC       1-PHOSPHATE = PYROPHOSPHATE + UDP-N-ACETYL-D-GLUCOSAMINE.
CC   -!- PATHWAY: PEPTIDOGLYCAN AND LIPOPOLYSACCHARIDE BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF
CC       ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT THAT CREATES TWO ORFS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X01631; CAA25784.1; -.
DR   EMBL; L10328; AAA62082.1; ALT_FRAME.
DR   EMBL; L10328; AAA62081.1; ALT_FRAME.
DR   EMBL; AE000450; AAC76753.1; -.
DR   ECOGENE; EG11198; GLMU.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; 1.
DR   PFAM; PF00132; hexapep; 3.
DR   PFAM; PF00483; NTP_transferase; 1.
KW   Peptidoglycan synthesis; Cell wall; Transferase;
KW   Nucleotidyltransferase; Repeat; Multifunctional enzyme.
FT   CONFLICT    186    187       KL -> NV (IN REF. 1).
SQ   SEQUENCE   456 AA;  49190 MW;  B9E65439 CRC32;

  Query Match           78.6%;  Score 44;  DB 1;  Length 456;
  Best Local Similarity  75.0%;  Pred. No. 1.52e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    124 GGIGLLTV 131
          :|||:|||
Qy      2 AGIGILTV 9


RESULT    8
ID   XYND_PAEPO     STANDARD;      PRT;   635 AA.
AC   P45796;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   ENDO-1,4-BETA-XYLANASE D PRECURSOR (EC 3.2.1.8) (XYLANASE D)
DE   (1,4-BETA-D-XYLAN XYLANOHYDROLASE D).
GN   XYND.
OS   Paenibacillus polymyxa (Bacillus polymyxa).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Paenibacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 842;
RX   MEDLINE; 92041687.
RA   GOSALBES M.J., PEREZ-GONZALEZ J.A., GONZALEZ R., NAVARRO A.;
RT   "Two beta-glycanase genes are clustered in Bacillus polymyxa:
RT   molecular cloning, expression, and sequence analysis of genes
RT   encoding a xylanase and an endo-beta-(1,3)-(1,4)-glucanase.";
RL   J. Bacteriol. 173:7705-7710(1991).
CC   -!- FUNCTION: SHOWS XYLANASE ACTIVITY AS WELL AS ALPHA-L-
CC       ARABINOFURANOSIDASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC       LINKAGES IN XYLANS.
CC   -!- PATHWAY: XYLAN DEGRADATION.
CC   -!- SIMILARITY: BELONGS TO FAMILY 43 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57094; CAA40378.1; -.
KW   Xylan degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    635       ENDO-1,4-BETA-XYLANASE D.
SQ   SEQUENCE   635 AA;  67914 MW;  078AAB82 CRC32;

  Query Match           78.6%;  Score 44;  DB 1;  Length 635;
  Best Local Similarity  75.0%;  Pred. No. 1.52e+01;
  Matches     6;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db    149 GAGIGVLT 156
          :||||:||
Qy      1 AAGIGILT 8


RESULT    9
ID   ATPE_ARATH     STANDARD;      PRT;   132 AA.
AC   P09468;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   ATP SYNTHASE EPSILON CHAIN (EC 3.6.1.34).
GN   ATPE.
OS   Arabidopsis thaliana (Mouse-ear cress).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
OC   euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC   core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC   Arabidopsis.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA;
RX   MEDLINE; 89057486.
RA   CHEN H.-C., WINTZ H., WEIL J.-H., PILLAY D.T.N.;
RT   "Nucleotide sequence of chloroplast CF1-ATPase epsilon-subunit and
RT   elongator tRNAMet genes from Arabidopsis thaliana.";
RL   Nucleic Acids Res. 16:10372-10372(1988).
CC   -!- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC       GRADIENT ACROSS THE MEMBRANE.
CC   -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC       CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC       SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC       HAS THREE MAIN SUBUNITS: A, B AND C.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE ATPASE EPSILON CHAIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12889; CAA31381.1; -.
DR   PIR; S01903; S01903.
DR   MENDEL; 13518; ARAth;atpE;2.
DR   PFAM; PF00401; ATP-synt_DE; 1.
KW   ATP synthesis; Chloroplast; Thylakoid membrane; CF(1);
KW   Hydrolase; Hydrogen ion transport.
SQ   SEQUENCE   132 AA;  14472 MW;  D826F274 CRC32;

  Query Match           76.8%;  Score 43;  DB 1;  Length 132;
  Best Local Similarity  66.7%;  Pred. No. 2.53e+01;
  Matches     6;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     43 AVDIGILTI 51
          |  |||||:
Qy      1 AAGIGILTV 9


RESULT   10
ID   SGAA_HYPME     STANDARD;      PRT;   404 AA.
AC   O08374;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   SERINE--GLYOXYLATE AMINOTRANSFERASE (EC 2.6.1.45) (SGAT).
GN   SGAA.
OS   Hyphomicrobium methylovorum.
OC   Bacteria; Proteobacteria; alpha subdivision; Hyphomicrobium.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GM2;
RX   MEDLINE; 97054583.
RA   HAGISHITA T., YOSHIDA T., IZUMI Y., MITSUNAGA T.;
RT   "Cloning and expression of the gene for serine-glyoxylate
RT   aminotransferase from an obligate methylotroph Hyphomicrobium
RT   methylovorum GM2.";
RL   Eur. J. Biochem. 241:1-5(1996).
CC   -!- CATALYTIC ACTIVITY: L-SERINE + GLYOXYLATE = 3-HYDROXYPYRUVATE +
CC       GLYCINE.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: SERINE PATHWAY.
CC   -!- SIMILARITY: BELONGS TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86125; BAA19919.1; -.
DR   PROSITE; PS00595; AA_TRANSFER_CLASS_5; FALSE_NEG.
DR   PFAM; PF00266; aminotran_5; 1.
KW   Transferase; Aminotransferase; Pyridoxal phosphate.
FT   INIT_MET      0      0
FT   BINDING     195    195       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   404 AA;  43762 MW;  8A1CAD51 CRC32;

  Query Match           76.8%;  Score 43;  DB 1;  Length 404;
  Best Local Similarity  55.6%;  Pred. No. 2.53e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db    200 PTGLGILAV 208
          ::|:|||:|
Qy      1 AAGIGILTV 9


RESULT   11
ID   YXCC_BACSU     STANDARD;      PRT;   461 AA.
AC   P46333; O32289;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN IOLS-HTPG INTERGENIC
DE   REGION.
GN   YXCC OR SS92BR.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / BGSC1A1;
RX   MEDLINE; 96093926.
RA   YOSHIDA K.-I., SEKI S., FUJIMURA M., MIWA Y., FUJITA Y.;
RT   "Cloning and sequencing of a 36-kb region of the Bacillus subtilis
RT   genome between the gnt and iol operons.";
RL   DNA Res. 2:61-69(1995).
RN   [2]
RP   REVISIONS.
RA   FUJITA Y., SHIBAYAMA T., ISHIO I., AOYAMA D., YOSHIDA K.-I.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB005554; BAA21604.1; -.
DR   EMBL; Z99124; CAB16017.1; -.
DR   SUBTILIST; BG11360; YXCC.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR   PFAM; PF00083; sugar_tr; 1.
KW   Hypothetical protein; Transport; Transmembrane.
FT   TRANSMEM     15     35       POTENTIAL.
FT   TRANSMEM     39     59       POTENTIAL.
FT   TRANSMEM     77     97       POTENTIAL.
```

```
FT   TRANSMEM     105    125       POTENTIAL.
FT   TRANSMEM     140    160       POTENTIAL.
FT   TRANSMEM     164    184       POTENTIAL.
FT   TRANSMEM     242    262       POTENTIAL.
FT   TRANSMEM     281    301       POTENTIAL.
FT   TRANSMEM     309    329       POTENTIAL.
FT   TRANSMEM     342    362       POTENTIAL.
FT   CONFLICT     400    401       RP -> SA (IN REF. 3).
SQ   SEQUENCE   461 AA;  50235 MW;  D47572C9 CRC32;

  Query Match           76.8%;  Score 43;  DB 1;  Length 461;
  Best Local Similarity  85.7%;  Pred. No. 2.53e+01;
  Matches     6;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Db    285 GIGILNV 291
          ||||| |
Qy      3 GIGILTV 9


RESULT   12
ID   ACHE_MOUSE     STANDARD;      PRT;   493 AA.
AC   P20782;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   ACETYLCHOLINE RECEPTOR PROTEIN, EPSILON CHAIN PRECURSOR.
GN   CHRNE OR ACRE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89214211.
RA   BUONANNO A., MUDD J., MERLIE J.P.;
RT   "Isolation and characterization of the beta and epsilon subunit genes
RT   of mouse muscle acetylcholine receptor.";
RL   J. Biol. Chem. 264:7611-7616(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91067487.
RA   GARDNER P.D.;
RT   "Nucleotide sequence of the epsilon-subunit of the mouse muscle
RT   nicotinic acetylcholine receptor.";
RL   Nucleic Acids Res. 18:6714-6714(1990).
CC   -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
CC       EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC       LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
CC       MEMBRANE.
CC   -!- SUBUNIT: PENTAMER OF TWO ALPHA CHAINS, AND ONE EACH OF THE BETA,
CC       DELTA, AND GAMMA (IN IMMATURE MUSCLE) OR EPSILON (IN MATURE
CC       MUSCLE) CHAINS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04698; AAA37153.1; -.
DR   EMBL; X55718; CAA39251.1; -.
DR   PIR; S13592; ACMSE.
DR   PIR; B33358; B33358.
DR   MGD; MGI:87894; ACRE.
DR   PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
DR   PFAM; PF00065; neur_chan; 1.
KW   Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
KW   Transmembrane.
FT   SIGNAL        1     20
FT   CHAIN        21    493       ACETYLCHOLINE RECEPTOR PROTEIN, EPSILON
FT                                 CHAIN.
FT   DOMAIN        21    239       EXTRACELLULAR.
FT   TRANSMEM     240    264       POTENTIAL.
FT   DOMAIN       265    272       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     273    291       POTENTIAL.
FT   DOMAIN       292    306       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     307    328       POTENTIAL.
FT   DOMAIN       329    456       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     457    480       POTENTIAL.
FT   DOMAIN       481    493       EXTRACELLULAR (POTENTIAL).
FT   DISULFID     148    162       BY SIMILARITY.
FT   CARBOHYD      86     86       POTENTIAL.
FT   CARBOHYD     161    161       PROBABLE.
SQ   SEQUENCE   493 AA;  54914 MW;  BB5BF2C0 CRC32;

  Query Match           76.8%;  Score 43;  DB 1;  Length 493;
  Best Local Similarity  75.0%;  Pred. No. 2.53e+01;
  Matches     6;  Conservative    1;  Mismatches   1;  Indels   0;  Gaps   0;

Db     93 AGVGILRV 100
          ||:||| |
Qy      2 AGIGILTV 9


RESULT   13
ID   YD3M_HERAU     STANDARD;      PRT;   611 AA.
AC   P25280;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   HYPOTHETICAL 68.4 KD PROTEIN IN HGIDIIM 3'REGION (ORF68).
OS   Herpetosiphon aurantiacus (Herpetosiphon giganteus).
OC   Bacteria; Green non-sulfur bacteria; Chloroflexaceae group;
OC   Herpetosiphon.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HPA2;
RX   MEDLINE; 92039068.
RA   DUESTERHOEFT A., KROEGER M.;
RT   "Cloning, sequence and characterization of m5C-methyltransferase-
RT   encoding gene, hgiDIIM (GTCGAC), from Herpetosiphon giganteus strain
RT   Hpa2.";
RL   Gene 106:87-92(1991).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55141; CAA38942.1; -.
DR   PIR; JT0592; JT0592.
DR   PIR; S21953; S21953.
DR   PIR; S21950; S21950.
KW   Hypothetical protein; Restriction system; Repeat.
FT   DOMAIN       382    403       2.5 X 11 AA TANDEM REPEATS.
FT   REPEAT       382    392       1.
FT   REPEAT       393    403       2.
FT   REPEAT       404    409       3 (INCOMPLETE).
SQ   SEQUENCE   611 AA;  68354 MW;  473CD6A4 CRC32;

  Query Match           76.8%;  Score 43;  DB 1;  Length 611;
  Best Local Similarity  71.4%;  Pred. No. 2.53e+01;
  Matches     5;  Conservative    2;  Mismatches   0;  Indels   0;  Gaps   0;

Db     97 GIGILAI 103
          |||||::
Qy      3 GIGILTV 9


RESULT   14
```

```
ID   VCAD_LAMBD     STANDARD;      PRT;   110 AA.
AC   P03712;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   HEAD DECORATION PROTEIN (GPD) (MAJOR CAPSID PROTEIN D).
GN   D.
OS   Bacteriophage lambda.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC   Lambda phage group.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 83189071.
RA   SANGER F., COULSON A.R., HONG G.F., HILL D.F., PETERSEN G.B.;
RT   "Nucleotide sequence of bacteriophage lambda DNA.";
RL   J. Mol. Biol. 162:729-773(1982).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE; 84207913.
RA   WITKIEWICZ H., SCHWEIGER M.;
RT   "The head protein D of bacterial virus lambda is related to
RT   eukaryotic chromosomal proteins.";
RL   EMBO J. 1:1559-1564(1982).
CC   -!- FUNCTION: STABILIZES THE HEAD SHELL FOLLOWING THE REARRANGEMENT
CC       OF THE GPE SUBUNITS OF THE HEAD SHELL LATTICE THAT ACCOMPANIES
CC       EXPANSION OF THE HEAD. THERE ARE APPROXIMATELY 420 COPIES OF
CC       PROTEIN D PER MATURE PHAGE.
CC   -!- SIMILARITY: TO BACTERIOPHAGE 21 HEAD DECORATION PROTEIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02459; AAA96539.1; -.
DR   PIR; A04334; VHBPDL.
DR   PIR; A23206; A23206.
KW   Coat protein.
SQ   SEQUENCE   110 AA;  11572 MW;  FDD50011 CRC32;

  Query Match           75.0%;  Score 42;  DB 1;  Length 110;
  Best Local Similarity  55.6%;  Pred. No. 4.15e+01;
  Matches     5;  Conservative    4;  Mismatches   0;  Indels   0;  Gaps   0;

Db     56 GAAVGILAV 64
          :|::|||:|
Qy      1 AAGIGILTV 9


RESULT   15
ID   FLA2_METVO     STANDARD;      PRT;   216 AA.
AC   P27804; P17602;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-DEC-1999 (Rel. 39, Last annotation update)
DE   FLAGELLIN B2 PRECURSOR.
GN   FLAB2.
OS   Methanococcus voltae.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PS;
RX   MEDLINE; 92041608.
RA   KALMOKOFF M.L., JARRELL K.F.;
RT   "Cloning and sequencing of a multigene family encoding the flagellins
RT   of Methanococcus voltae.";
RL   J. Bacteriol. 173:7113-7125(1991).
RN   [2]
RP   SEQUENCE OF 13-32.
RC   STRAIN=PS;
RX   MEDLINE; 90179742.
RA   KALMOKOFF M.L., KARNAUCHOW T.M., JARRELL K.F.;
RT   "Conserved N-terminal sequences in the flagellins of archaebacteria.";
RL   Biochem. Biophys. Res. Commun. 167:154-160(1990).
CC   -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC       FORM THE FILAMENTS OF FLAGELLA.
CC   -!- SIMILARITY: BELONGS TO THE ARCHAEBACTERIAL FLAGELLIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M72148; AAA73075.1; -.
DR   PIR; A34624; A34624.
DR   PIR; C41316; C41316.
KW   Flagella; Multigene family.
FT   PROPEP          1      12
FT   CHAIN          13     216       FLAGELLIN B2.
SQ   SEQUENCE   216 AA;  22799 MW;  B62A0B23 CRC32;

  Query Match           75.0%;  Score 42;  DB 1;  Length 216;
  Best Local Similarity  66.7%;  Pred. No. 4.15e+01;
  Matches     6;  Conservative    1;  Mismatches   2;  Indels   0;  Gaps   0;

Db     13 ASGIGTLIV 21
          |:||| | |
Qy      1 AAGIGILTV 9


Search completed: Fri May  5 22:01:33 2000
Job time : 40 secs.
```

```
   #cross-references MUID:93076110
   #accession    JQ1674
      ##molecule_type DNA
      ##residues       1-942 ##label CHA
      ##cross-references GB:L00670; NID:g166887; PIDN:AAA32876.1; PID:g166888
CLASSIFICATION   #superfamily protein kinase Xa21; leucine-rich
                   alpha-2-glycoprotein repeat homology; protein kinase
                   homology
KEYWORDS         ATP; autophosphorylation; glycoprotein; phosphotransferase;
                   receptor; serine/threonine-specific protein kinase; tandem
                   repeat; transmembrane protein
FEATURE
   1-22              #domain signal sequence #status predicted #label SIG\
   23-942            #product protein kinase TMK1 #status predicted #label
                       MAT\
   65-88             #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR1\
   89-111            #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR2\
   112-135           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR3\
   136-160           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR4\
   161-186           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR5\
   187-209           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR6\
   210-232           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR7\
   233-255           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR8\
   256-279           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LRR9\
   280-299           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR10\
   300-323           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR11\
   324-346           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR12\
   363-386           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR14\
   387-410           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR15\
   411-434           #domain leucine-rich alpha-2-glycoprotein repeat
                       homology #label LR16\
   480-503           #domain transmembrane #status predicted #label TMM\
   586-872           #domain protein kinase homology #label KIN\
   594-602           #region protein kinase ATP-binding motif\
   86,99,158,164,171,
   363,533,587       #binding_site carbohydrate (Asn) (covalent) #status
                       predicted\
   616,634,717,719   #active_site Lys, Glu, Asp, Lys #status predicted
SUMMARY          #length 942  #molecular-weight 102387  #checksum 2851

  Query Match             9.9%;  Score 88;  DB 1;  Length 942;
  Best Local Similarity  37.5%;  Pred. No. 8.34e-01;
  Matches     9;  Conservative   10;  Mismatches   4;  Indels   1;  Gaps   1;

Db    491 GLLSIFL-IGLLVFCWYKKRQKRF 513
          |:|:::| : ||: ||| :|:: :
Qy     31 GILTVILGVLLLIGCWYCRRRNGY 54


RESULT   14
ENTRY            F71010     #type complete
TITLE            hypothetical protein PH1380 - Pyrococcus horikoshii
ORGANISM         #formal_name Pyrococcus horikoshii
DATE             14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
                   14-Aug-1998
ACCESSIONS       F71010
REFERENCE        A71000
   #authors      Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
                   Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                   Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                   Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
                   Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
                   A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                   Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
   #journal      DNA Res. (1998) 5:55-76
   #title        Complete sequence and gene organization of the genome of a
                   hyper-thermophilic archaebacterium, Pyrococcus horikoshii
                   OT3.
   #cross-references MUID:98344137
   #accession    F71010
      ##status        preliminary; nucleic acid sequence not shown;
                        translation not shown
      ##molecule_type DNA
      ##residues      1-442 ##label KAW
      ##cross-references GB:AP000006; NID:g3236133; PID:d1031429; PID:g3257803
      ##experimental_source strain OT3
      ##note          this accession replaces an interim accession for a
                        sequence replaced by GenBank
GENETICS
   #gene         PH1380
SUMMARY          #length 442  #molecular-weight 49017  #checksum 3898

  Query Match             9.8%;  Score 87;  DB 2;  Length 442;
  Best Local Similarity  56.3%;  Pred. No. 1.16e+00;
  Matches     9;  Conservative    6;  Mismatches   1;  Indels   0;  Gaps   0;

Db     95 PGVGILVIFLGMLLLV 110
          :|:||| ::||:|||:
Qy     28 AGIGILTVILGVLLLI 43


RESULT   15
ENTRY            YRHUR2     #type complete
TITLE            dopachrome Delta-isomerase (EC 5.3.3.12) precursor - human
ALTERNATE_NAMES  dopachrome conversion factor; dopachrome oxidoreductase;
                   dopachrome tautomerase; tyrosinase-related protein 2
ORGANISM         #formal_name Homo sapiens #common_name man
DATE             13-Jan-1995 #sequence_revision 17-May-1996 #text_change
                   11-Jun-1999
ACCESSIONS       S43510; S41089; I53786
REFERENCE        S43510
   #authors      Yokoyama, K.; Suzuki, H.; Yasumoto, K.; Tomita, Y.;
                   Shibahara, S.
   #journal      Biochim. Biophys. Acta (1994) 1217:317-321
   #title        Molecular cloning and functional analysis of a cDNA coding
                   for human DOPAchrome tautomerase/tyrosinase-related
                   protein-2.
   #cross-references MUID:94198295
   #accession    S43510
      ##molecule_type mRNA
      ##residues      1-519 ##label YOK
      ##cross-references EMBL:D17547; NID:g484259; PIDN:BAA04484.1;
                          PID:d1005009; PID:g914938
REFERENCE        S41089
   #authors      Bouchard, B.; del Marmol, V.; Jackson, I.J.; Cherif, D.;
                   Dubertret, L.
   #journal      Eur. J. Biochem. (1994) 219:127-134
   #title        Molecular characterization of a human
                   tyrosinase-related-protein-2 cDNA. Patterns of expression
                   in melanocytic cells.
   #cross-references MUID:94139684
   #accession    S41089
      ##molecule_type mRNA
      ##residues      1-519 ##label BOU
      ##cross-references GB:S69231; NID:g545618; PIDN:AAC60627.1; PID:g545619
REFERENCE        A44749
   #authors      Pawelek, J.M.
   #journal      Biochem. Biophys. Res. Commun. (1990) 166:1328-1333
   #title        Dopachrome conversion factor as an isomerase.
   #cross-references MUID:90165938
   #contents     annotation; reaction description
REFERENCE        I53786
```

```
   #authors      Cassady, J.L.; Sturm, R.A.
   #journal      Gene (1994) 143:295-298
   #title        Sequence of the human dopachrome tautomerase-encoding TRP-2
                   cDNA.
   #cross-references MUID:94266170
   #accession    I53786
     ##status        translated from GB/EMBL/DDBJ
     ##molecule_type mRNA
     ##residues      1-519 ##label RES
     ##cross-references GB:L18967; NID:g399581; PIDN:AAA20870.1; PID:g399582
     ##experimental_source melanoma cell line A2058
GENETICS
   #gene         GDB:DCT; TYRP2
     ##cross-references GDB:231628; OMIM:191275
   #map_position 13q32-13q32
FUNCTION
   #description  catalyzes the isomerization between 2-carboxy-1,2,3,
                   5-tetrahydroindolinium (dopachrome tautomer) and 5,
                   6-dihydroxyindole-2-carboxylic acid
   #pathway      melanin biosynthesis
CLASSIFICATION   #superfamily monophenol monooxygenase
KEYWORDS         copper; glycoprotein; intramolecular oxidoreductase;
                   isomerase; melanin biosynthesis; transmembrane protein
FEATURE
   1-23              #domain signal sequence #status predicted #label SIG\
   24-519            #product dopachrome Delta-isomerase #status predicted
                       #label MAT\
   475-493           #domain transmembrane #status predicted #label TRM\
   170,237,300,342,
   377               #binding_site carbohydrate (Asn) (covalent) #status
                       predicted\
   189,211,220,223   #binding_site copper (His) #status predicted\
   369,373,396,426   #binding_site copper (His) #status predicted
SUMMARY          #length 519  #molecular-weight 59145  #checksum 989

  Query Match             9.8%;  Score 87;  DB 1;  Length 519;
  Best Local Similarity  35.3%;  Pred. No. 1.16e+00;
  Matches    12;  Conservative   11;  Mismatches  10;  Indels   1;  Gaps   1;

Db    477 MGTLVALVGLFVLLAFLQYRRLRKGYTPLMETHL 510
          :| |  ::|:::|::    || |:|| :||:  |
Qy     30 IGILTVILGVLLLIGCWYCRR-RNGYRALMDKSL 62


Search completed: Fri May  5 21:46:28 2000
Job time : 111 secs.
```

```
                  Sugano, S.
   #journal       Genomics (1998) 49:458-461
   #title         Cloning, expression analysis, and chromosomal localization of
                    BH-protocadherin (PCDH7), a novel member of the cadherin
                    superfamily.
   #cross-references MUID:98277460
   #accession     T00041
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1072 ##label YOS
      ##cross-references EMBL:AB006756; NID:d1184678; PID:d1026123
      ##experimental_source clone BH-Pcdh-b
GENETICS
   #map_position 4p15
SUMMARY           #length 1072  #molecular-weight 116462  #checksum 9727

  Query Match           11.1%;  Score 99;  DB 2;  Length 1072;
  Best Local Similarity  50.0%;  Pred. No. 1.88e-02;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    887 GIMTVILIILIVVMARYCRSKNKNGYEA 914
          ||:|||| :|::: :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    6
ENTRY             T00042     #type complete
TITLE             BH-protocadherin PCDH7 (clone BH-Pcdh-c) - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                    07-May-1999
ACCESSIONS        T00042
REFERENCE         Z14074
   #authors       Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                    Sugano, S.
   #journal       Genomics (1998) 49:458-461
   #title         Cloning, expression analysis, and chromosomal localization of
                    BH-protocadherin (PCDH7), a novel member of the cadherin
                    superfamily.
   #cross-references MUID:98277460
   #accession     T00042
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1200 ##label YOS
      ##cross-references EMBL:AB006757; NID:d1184679; PID:d1026124
      ##experimental_source clone BH-Pcdh-c
GENETICS
   #map_position 4p15
SUMMARY           #length 1200  #molecular-weight 130337  #checksum 7152

  Query Match           11.1%;  Score 99;  DB 2;  Length 1200;
  Best Local Similarity  50.0%;  Pred. No. 1.88e-02;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    840 GIMTVILIILIVVMARYCRSKNKNGYEA 867
          ||:|||| :|::: :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    7
ENTRY             C71600     #type complete
TITLE             rifin PFB1020w - malaria parasite (Plasmodium falciparum)
ORGANISM          #formal_name Plasmodium falciparum
DATE              13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change
                    07-May-1999
ACCESSIONS        C71600
REFERENCE         A71600
   #authors       Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.;
                    Aravind, L.; Koonin, E.V.; Shallom, S.; Mason, T.; Yu, K.;
                    Fujii, C.; Pederson, J.; Shen, K.; Jing, J.; Aston, C.;
                    Lai, Z.; Schwartz, D.C.; Pertea, M.; Salzberg, S.; Zhou,
                    L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
                    Fraser, C.M.; Adams, M.D.; Venter, J.C.; Hoffman, S.L.
   #journal       Science (1998) 282:1126-1132
   #title         Chromosome 2 sequence of the human malaria parasite
                    Plasmodium falciparum.
   #cross-references MUID:99021743
   #accession     C71600
      ##status         preliminary; nucleic acid sequence not shown;
                         translation not shown
      ##molecule_type DNA
      ##residues       1-304 ##label GAR
      ##cross-references GB:AE001433; GB:AE001362; NID:g3845336; PID:g3845337;
                         TIGR:PFB1020w
      ##experimental_source clone 3D7
GENETICS
   #gene          PFB1020w
SUMMARY           #length 304  #molecular-weight 34000  #checksum 4789

  Query Match           10.8%;  Score 96;  DB 2;  Length 304;
  Best Local Similarity  45.5%;  Pred. No. 5.44e-02;
  Matches    15;  Conservative    9;  Mismatches   6;  Indels   3;  Gaps   3;

Db    263 EPCGIAALVLLILAVVLIILYIWLYRRRKNSYK 295
          |: ||: | : ||:|:|:|   | | ||:|:|:
Qy     26 EAAGIGILTV-ILGVLLLIGC-W-YCRRRNGYR 55


RESULT    8
ENTRY             T00043     #type complete
TITLE             BH-protocadherin-a - mouse
ORGANISM          #formal_name Mus musculus #common_name house mouse
DATE              22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                    22-Jan-1999
ACCESSIONS        T00043
REFERENCE         Z14075
   #authors       Yoshida, K.
   #submission    submitted to the EMBL Data Library, August 1997
   #accession     T00043
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-1069 ##label YOS
      ##cross-references EMBL:AB006758; NID:d1227200; PID:d1033562
GENETICS
   #gene          pcdh7
   #map_position 5C3-D
SUMMARY           #length 1069  #molecular-weight 116313  #checksum 4821

  Query Match           10.8%;  Score 96;  DB 2;  Length 1069;
  Best Local Similarity  50.0%;  Pred. No. 5.44e-02;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    887 GIMTVILIILIVMMARYCRSKNKNGYEA 914
          ||:|||| :|::: :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    9
ENTRY             I37202     #type complete
TITLE             B-CAM protein - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
                    28-Feb-1997
ACCESSIONS        I37202; S47272
REFERENCE         I37202
   #authors       Campbell, I.G.; Foulkes, W.D.; Senger, G.; Trowsdale, J.;
                    Garin-Chesa, P.; Rettig, W.J.
   #journal       Cancer Res. (1994) 54:5761-5765
   #title         Molecular cloning of the B-CAM cell surface glycoprotein of
                    epithelial cancers: a novel member of the immunoglobulin
                    superfamily.
   #cross-references MUID:95042297
   #accession     I37202
      ##status         preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues       1-588 ##label RES
```

```
     ##cross-references EMBL:X80026; NID:g535178; PID:g535179
GENETICS
   #gene          B-CAM
SUMMARY           #length 588  #molecular-weight 63566  #checksum 255

  Query Match           10.1%;  Score 90;  DB 2;  Length 588;
  Best Local Similarity  34.4%;  Pred. No. 4.28e-01;
  Matches    11;  Conservative   12;  Mismatches   8;  Indels   1;  Gaps   1;

Db    544 TSQAGVAVMAVAVSVGLLLLVVAVFYCVRRKG 575
          |:: : ::  ::| :|:|||: :  || ||:|
Qy     23 TAEEAAGIGILTVILGVLLLI-GCWYCRRRNG 53


RESULT   10
ENTRY             I38000     #type complete
TITLE             Lutheran blood group glycoprotein precursor - human
ORGANISM          #formal_name Homo sapiens #common_name man
DATE              09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change
                    29-Aug-1997
ACCESSIONS        I38000; S51663
REFERENCE         I38000
   #authors       Parsons, S.F.; Mallinson, G.; Holmes, C.H.; Houlihan, J.M.;
                    Simpson, K.L.; Mawby, W.J.; Spurr, N.K.; Warne, D.;
                    Barclay, A.N.; Anstee, D.J.
   #journal       Proc. Natl. Acad. Sci. U.S.A. (1995) 92:5496-5500
   #title         The Lutheran blood group glycoprotein, another member of the
                    immunoglobulin superfamily, is widely expressed in human
                    tissues and is developmentally regulated in human liver.
   #cross-references MUID:95296337
   #accession     I38000
      ##molecule_type mRNA
      ##residues       1-628 ##label RES
      ##cross-references EMBL:X83425; NID:g603559; PID:g603560
      ##note          parts of this sequence, including the amino end of the
                       mature form, were confirmed by peptide sequencing
GENETICS
   #gene          GDB:LU
      ##cross-references GDB:120155; OMIM:111200
   #map_position 19q12-19q13
KEYWORDS          glycoprotein
FEATURE
   1-31               #domain signal sequence #status predicted #label SIG\
   32-628             #product Lutheran blood group glycoprotein #status
                        experimental #label MAT
SUMMARY           #length 628  #molecular-weight 67374  #checksum 416

  Query Match           10.1%;  Score 90;  DB 2;  Length 628;
  Best Local Similarity  34.4%;  Pred. No. 4.28e-01;
  Matches    11;  Conservative   12;  Mismatches   8;  Indels   1;  Gaps   1;

Db    544 TSQAGVAVMAVAVSVGLLLLVVAVFYCVRRKG 575
          |:: : ::  ::| :|:|||: :  || ||:|
Qy     23 TAEEAAGIGILTVILGVLLLI-GCWYCRRRNG 53


RESULT   11
ENTRY             D72109     #type complete
TITLE             hypothetical protein - Chlamydia pneumoniae (strain CWL029)
ORGANISM          #formal_name Chlamydia pneumoniae
DATE              23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
                    23-Apr-1999
ACCESSIONS        D72109
REFERENCE         A72000
   #authors       Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.;
                    Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
   #journal       Nature Genet. (1999) 21:385-389
   #title         Comparative genomes of Clamydia pneumoniae and C.
                    trachomatis.
   #cross-references MUID:99206606
   #accession     D72109
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-91 ##label ARN
      ##cross-references GB:AE001604; GB:AE001363; NID:g4376438; PID:g4376442
      ##experimental_source strain CWL029
GENETICS
   #gene          CPn0173
SUMMARY           #length 91  #molecular-weight 10192  #checksum 1395

  Query Match            9.9%;  Score 88;  DB 2;  Length 91;
  Best Local Similarity  45.7%;  Pred. No. 8.34e-01;
  Matches    16;  Conservative    9;  Mismatches   6;  Indels   4;  Gaps   3;

Db     55 YGTSGETASLVFGILSLIVLVLLLIIECRNRECCRR 89
          | |: |:|: : |||::|: |||:| |:   | ||
Qy     21 YTTAEEAAG-I-GILTVILGVLLLIGCWY--CRRR 51


RESULT   12
ENTRY             F71923     #type complete
TITLE             hypothetical protein jhp0512 - Helicobacter pylori (strain
                    J99)
ORGANISM          #formal_name Helicobacter pylori
   #variety       strain J99
DATE              12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
                    29-Sep-1999
ACCESSIONS        F71923
REFERENCE         A71800
   #authors       Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
                    Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
                    B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
                    Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
                    Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
                    G.F.; Trust, T.J.
   #journal       Nature (1999) 397:176-180
   #title         Genomic sequence comparison of two unrelated isolates of the
                    human gastric pathogen Helicobacter pylori.
   #cross-references MUID:99120557
   #accession     F71923
      ##status        preliminary
      ##molecule_type DNA
      ##residues      1-215 ##label ARN
      ##cross-references GB:AE001484; GB:AE001439; NID:g4155043;
                          PIDN:AAD06088.1; PID:g4155051
      ##experimental_source strain J99
GENETICS
   #gene          jhp0512
CLASSIFICATION    #superfamily conserved hypothetical protein HI0219
SUMMARY           #length 215  #molecular-weight 23893  #checksum 1978

  Query Match            9.9%;  Score 88;  DB 2;  Length 215;
  Best Local Similarity  55.0%;  Pred. No. 8.34e-01;
  Matches    11;  Conservative    5;  Mismatches   4;  Indels   0;  Gaps   0;

Db    104 AEGLGITIMILGILVLLGLW 123
          | |:||  :|||:|:|:| |
Qy     27 AAGIGILTVILGVLLLIGCW 46


RESULT   13
ENTRY             JQ1674     #type complete
TITLE             protein kinase TMK1 (EC 2.7.1.-), receptor type precursor -
                    Arabidopsis thaliana
ORGANISM          #formal_name Arabidopsis thaliana #common_name mouse-ear
                    cress
DATE              30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
                    11-Jun-1999
ACCESSIONS        JQ1674
REFERENCE         JQ1674
   #authors       Chang, C.; Schaller, G.E.; Patterson, S.E.; Kwok, S.F.;
                    Meyerowitz, E.M.; Bleecker, A.B.
   #journal       Plant Cell (1992) 4:1263-1271
   #title         The TMK1 gene from Arabidopsis codes for a protein with
                    structural and biochemical characteristics of a receptor
                    protein kinase.
```

****

MPSRCH (TM)

****

Release 3.1A John F. Collins, Biocomputing Research Unit.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri May 5 21:44:37 2000; MasPar time 9.61 Seconds
579.049 Million cell updates/sec

Tabular output not generated.

Title: >US-09-267-439-2
Description: (1-118) from US09267439.pep
Perfect Score: 889
Sequence: 1 MPREDAHFIYGYPKKGHGHS.....NAPPAYEKLSAEQSPPPYSP 118

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 40.345; Variance 70.060; scale 0.576

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	889	100.0	118	2	A55253	melanoma antigen MART	5.10e-179
2	99	11.1	344	2	I49585	CD2 antigen protein p	1.88e-02
3	99	11.1	344	2	B28967	T-cell surface glycop	1.88e-02
4	99	11.1	1069	2	T00040	BH-protocadherin PCDH	1.88e-02
5	99	11.1	1072	2	T00041	BH-protocadherin PCDH	1.88e-02
6	99	11.1	1200	2	T00042	BH-protocadherin PCDH	1.88e-02
7	96	10.8	304	2	C71600	rifin PFB1020w - mala	5.44e-02
8	96	10.8	1069	2	T00043	BH-protocadherin-a -	5.44e-02
9	90	10.1	588	2	I37202	B-CAM protein - human	4.28e-01
10	90	10.1	628	2	I38000	Lutheran blood group	4.28e-01
11	88	9.9	91	2	D72109	hypothetical protein	8.34e-01
12	88	9.9	215	2	F71923	hypothetical protein	8.34e-01
13	88	9.9	942	1	JQ1674	protein kinase TMK1 (	8.34e-01
14	87	9.8	442	2	F71010	hypothetical protein	1.16e+00
15	87	9.8	519	1	YRHUR2	dopachrome Delta-isom	1.16e+00
16	87	9.8	682	2	T04846	protein kinase homolo	1.16e+00
17	87	9.8	1557	2	D41214	protein-tyrosine-phos	1.16e+00
18	87	9.8	1630	2	C41214	protein-tyrosine-phos	1.16e+00
19	86	9.7	517	2	S19243	tyrosinase-related pr	1.61e+00
20	86	9.7	704	2	A48040	meprin A (EC 3.4.24.1	1.61e+00
21	86	9.7	774	2	JC6095	hepatocyte nuclear fa	1.61e+00
22	85	9.6	320	2	S32966	probable membrane pro	2.22e+00
23	85	9.6	918	1	PXBY1P	H+-transporting ATPas	2.22e+00
24	84	9.4	227	2	T05800	probable transcriptio	3.07e+00
25	84	9.4	327	1	HLHUCD	T-cell surface glycop	3.07e+00
26	84	9.4	334	2	E71706	hypothetical protein	3.07e+00
27	84	9.4	512	2	F71839	nadh oxidoreductase I	3.07e+00
28	83	9.3	207	2	H72056	holliday junction hel	4.22e+00
29	83	9.3	350	2	S75065	sensory transduction	4.22e+00
30	83	9.3	700	1	HYHUMB	meprin A (EC 3.4.24.1	4.22e+00
31	83	9.3	824	1	S50767	S-receptor kinase (EC	4.22e+00
32	82	9.2	133	2	E71833	ribosomal protein L15	5.79e+00
33	82	9.2	155	2	F72086	hypothetical protein	5.79e+00
34	82	9.2	215	2	E64590	hypothetical protein	5.79e+00
35	82	9.2	416	2	G69748	conserved hypothetica	5.79e+00
36	82	9.2	477	2	A34368	interferon gamma rece	5.79e+00
37	82	9.2	516	2	H64897	amino acid transport	5.79e+00
38	82	9.2	897	1	A39255	cytokine receptor com	5.79e+00
39	82	9.2	1010	2	S45389	probable membrane pro	5.79e+00
40	82	9.2	1015	2	JC5062	phogrin precursor - h	5.79e+00
41	82	9.2	1015	2	JC5263	transmembrane tyrosin	5.79e+00
42	82	9.2	1807	2	JC6319	integrin beta-4 chain	5.79e+00
43	81	9.1	457	2	T12399	NADH dehydrogenase su	7.91e+00
44	81	9.1	511	2	S44275	dopamine receptor pro	7.91e+00
45	81	9.1	774	2	A70010	NADH dehydrogenase ho	7.91e+00

ALIGNMENTS

RESULT 1
ENTRY A55253 #type complete
TITLE melanoma antigen MART-1 - human
ALTERNATE_NAMES melan-A protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 10-Sep-1997
ACCESSIONS A55253; I38506
REFERENCE A55253
#authors Kawakami, Y.; Eliyahu, S.; Delgado, C.H.; Robbins, P.F.; Rivoltini, L.; Topalian, S.L.; Miki, T.; Rosenberg, S.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:3515-3519
#title Cloning of the gene coding for a shared human melanoma antigen recognized by autologous T cells infiltrating into tumor.
#cross-references MUID:94224770
#accession A55253
##status preliminary
##molecule_type mRNA
##residues 1-118 ##label KAW
##cross-references GB:U06452; NID:g476131; PID:g476132
REFERENCE I38506
#authors Coulie, P.G.; Brichard, V.; Van Pel, A.; Wolfel, T.; Schneider, J.; Traversari, C.; Mattei, S.; De Plaen, E.; Lurquin, C.; Szikora, J.P.; Renauld, J.; Boon, T.
#journal J. Exp. Med. (1994) 180:35-42
#title A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas [see comments].
#cross-references MUID:94275389
#accession I38506
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues 1-118 ##label RES
##cross-references EMBL:U06654; NID:g517022; PID:g517023
GENETICS
#gene GDB:MLANA
##cross-references GDB:358979
#map_position 17q21-17q24
SUMMARY #length 118 #molecular-weight 13157 #checksum 3535

Query Match 100.0%; Score 889; DB 2; Length 118;
Best Local Similarity 100.0%; Pred. No. 5.10e-179;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Qy      1 MPREDAHFIYGYPKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDK 60

Db     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     61 SLHVGTQCALTRRCPQEGFDHRDSKVSLQEKNCEPVVPNAPPAYEKLSAEQSPPPYSP 118


RESULT    2
ENTRY           I49585      #type complete
TITLE           CD2 antigen protein precursor - mouse
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
                  23-Jul-1999
ACCESSIONS      I49585
REFERENCE       I49585
   #authors     Yagita, H.; Okumura, K.; Nakauchi, H.
   #journal     J. Immunol. (1988) 140:1321-1326
   #title       Molecular cloning of the murine homologue of CD2: Homology of
                  the molecule to its human counterpart T11.
   #cross-references MUID:88140313
   #accession   I49585
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-344 ##label RES
      ##cross-references GB:M18934; NID:g192486; PIDN:AAA37397.1; PID:g309158
CLASSIFICATION  #superfamily T-cell surface glycoprotein CD2
SUMMARY         #length 344  #molecular-weight 38337  #checksum 4699

  Query Match           11.1%;  Score 99;  DB 2;  Length 344;
  Best Local Similarity  40.4%;  Pred. No. 1.88e-02;
  Matches    21;  Conservative    9;  Mismatches  18;  Indels   4;  Gaps   4;

Db    198 PEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEELEI 245
          | || :  | |   :|| |:| |:| | |:| |  |:||:  |  :|  | :
Qy     13 PKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSLHV 64


RESULT    3
ENTRY           B28967      #type complete
TITLE           T-cell surface glycoprotein CD2 precursor - mouse
ALTERNATE_NAMES CD2 antigen; T-lymphocyte antigen CD2; T11 protein
ORGANISM        #formal_name Mus musculus #common_name house mouse
DATE            30-Jun-1989 #sequence_revision 03-Jun-1993 #text_change
                  23-Jul-1999
ACCESSIONS      B28967; S01347; S02293
REFERENCE       A28967
   #authors     Diamond, D.J.; Clayton, L.K.; Sayre, P.H.; Reinherz, E.L.
   #journal     Proc. Natl. Acad. Sci. U.S.A. (1988) 85:1615-1619
   #title       Exon-intron organization and sequence comparison of human and
                  murine T11 (CD2) genes.
   #cross-references MUID:88144486
   #accession   B28967
      ##molecule_type mRNA
      ##residues     1-344 ##label DIA
      ##cross-references GB:M19807; NID:g192479; PIDN:AAA37393.1; PID:g387122;
                        GB:J03622; GB:J03623
      ##note        the authors translated the codon TAT for residue 99 as
                     Thr
REFERENCE       S01347
   #authors     Clayton, L.K.; Sayre, P.H.; Novotny, J.; Reinherz, E.L.
   #journal     Eur. J. Immunol. (1987) 17:1367-1370
   #title       Murine and human T11 (CD2) cDNA sequences suggest a common
                  signal transduction mechanism.
   #cross-references MUID:88004738
   #accession   S01347
      ##molecule_type mRNA
      ##residues     1-127,'M',129-174,'N',176-190,'NM',193-344 ##label CLA
      ##cross-references EMBL:X06143; NID:g54223; PIDN:CAA29500.1; PID:g54224
REFERENCE       S02293
   #authors     Sewell, W.A.; Brown, M.H.; Owen, M.J.; Fink, P.J.; Kozak,
                  C.A.; Crumpton, M.J.
   #journal     Eur. J. Immunol. (1987) 17:1015-1020
   #title       The murine homologue of the T lymphocyte CD2 antigen:
                  molecular cloning, chromosome assignment and cell surface
                  expression.
   #cross-references MUID:87276135
   #accession   S02293
      ##status       not compared with conceptual translation
      ##molecule_type mRNA
      ##residues     1-127,'M',129-174,'N',176-191,'M',193-344 ##label SEW
      ##cross-references EMBL:Y00023; NID:g50346; PIDN:CAA68258.1; PID:g50347
GENETICS
   #map_position 3
CLASSIFICATION  #superfamily T-cell surface glycoprotein CD2
KEYWORDS        glycoprotein; surface antigen; T-cell; transmembrane protein
FEATURE
   1-22             #domain signal sequence #status predicted #label SIG\
   23-344           #product T-cell surface glycoprotein CD2 #status
                      predicted #label MAT\
   23-203           #domain extracellular #status predicted #label EXT\
   204-228          #domain transmembrane #status predicted #label TMM\
   229-344          #domain intracellular #status predicted #label INT
SUMMARY         #length 344  #molecular-weight 38325  #checksum 4974

  Query Match           11.1%;  Score 99;  DB 2;  Length 344;
  Best Local Similarity  40.4%;  Pred. No. 1.88e-02;
  Matches    21;  Conservative    9;  Mismatches  18;  Indels   4;  Gaps   4;

Db    198 PEKGLSF-YVTVGVGAG-GLLLVLL-VALFIFC-ICKRRKRNRRRKDEELEI 245
          | || :  | |   :|| |:| |:| | |:| |  |:||:  |  :|  | :
Qy     13 PKKGHGHSYTTAEEAAGIGILTVILGVLLLIGCWYCRRRNGYRALMDKSLHV 64


RESULT    4
ENTRY           T00040      #type complete
TITLE           BH-protocadherin PCDH7 - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                  07-May-1999
ACCESSIONS      T00040
REFERENCE       Z14074
   #authors     Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
                  Sugano, S.
   #journal     Genomics (1998) 49:458-461
   #title       Cloning, expression analysis, and chromosomal localization of
                  BH-protocadherin (PCDH7), a novel member of the cadherin
                  superfamily.
   #cross-references MUID:98277460
   #accession   T00040
      ##status       preliminary; translated from GB/EMBL/DDBJ
      ##molecule_type mRNA
      ##residues     1-1069 ##label YOS
      ##cross-references EMBL:AB006755; NID:d1184677; PID:d1026122
      ##experimental_source clone BH-Pcdh-a
GENETICS
   #map_position 4p15
SUMMARY         #length 1069  #molecular-weight 116104  #checksum 9974

  Query Match           11.1%;  Score 99;  DB 2;  Length 1069;
  Best Local Similarity  50.0%;  Pred. No. 1.88e-02;
  Matches    14;  Conservative    8;  Mismatches   4;  Indels   2;  Gaps   2;

Db    887 GIMTVILIILIVVMARYCRSKNKNGYEA 914
          ||:|||| :|:::  :||| : :||| |
Qy     31 GILTVILGVLLLIGCWYCR-R-RNGYRA 56


RESULT    5
ENTRY           T00041      #type complete
TITLE           BH-protocadherin PCDH7 (clone BH-Pcdh-b) - human
ORGANISM        #formal_name Homo sapiens #common_name man
DATE            22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change
                  07-May-1999
ACCESSIONS      T00041
REFERENCE       Z14074
   #authors     Yoshida, K.; Yoshitomo-Nakagawa, K.; Seki, N.; Sasaki, M.;
```